The Thesis Committee for Michael B. Cammarata
Certifies that this is the approved version of the following thesis:

Development of Top-Down Methods for Evaluating Protein Structure and Protein Unfolding Utilizing 193 nm Ultraviolet Photodissociation Mass Spectrometry

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Jennifer S. Brodbelt

Lauren J. Webb

Development of Top-Down Methods for Evaluating Protein Structure and Protein Unfolding Utilizing 193 nm Ultraviolet Photodissociation Mass Spectrometry

by

Michael B. Cammarata, B.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

December 2014

Acknowledgements

Funding from the NSF (CHE-1012622), (CHE-1402753) and the Welch Foundation (F-1155) are acknowledged.

Abstract

Development of Top-Down Methods for Evaluating Protein Structure and Protein Unfolding Utilizing 193 nm Ultraviolet Photodissociation Mass Spectrometry

Michael B. Cammarata, M.A.

The University of Texas at Austin, 2014

Supervisor: Jennifer Brodbelt

Ultraviolet photodissocation (UVPD) mass spectrometry was used for high mass accuracy top down characterization of two proteins labeled by the chemical probe, S-ethylacetimidate (SETA), in order to evaluate conformational changes as a function of denaturation. The SETA labeling/UVPD-MS methodology was used to monitor the mild denaturation of horse heart myoglobin by acetonitrile, and the results showed good agreement with known acetonitrile and acid unfolding pathways of myoglobin. UVPD outperformed another ion activation method, electron transfer dissociation (ETD), in terms of sequence coverage, allowing the SETA reactivity of greater number of lysine amines to be monitored and thus providing a more detailed map of myoglobin. This strategy was applied to the third zinc-finger binding domain, domain C, of PARP-1 (PARP-C), to evaluate the discrepancies between the NMR and crystal structures which reported monomer and dimer forms of the protein, respectively. The trends reflected from the reactivity of each lysine as a function of acetonitrile denaturation supported that PARP-C exists as a monomer in solution with a close-packed C-

terminal alpha helix. Additionally, those lysines for which the SETA reactivity increased under denaturing conditions were found to engage in tertiary polar contacts such as salt bridging and hydrogen bonding, providing evidence that the SETA/UVPD-MS approach offers a versatile means to probe the interactions responsible for conformational changes in proteins.

UVPD mass spectrometry was also employed to investigate the structure of holo-myoglobin as well as its apo form transferred to the gas phase by native electrospray. The fragmentation yields from UVPD showed the greatest overall correlation with B-factors generated from the crystal structure of apo-myoglobin, particularly for the more disordered loop regions. Comparison of UVPD of holo- and apo- myoglobin revealed similarities in fragmentation yields, particularly for the lower charge states (8 and 9+), but those regions involved in harboring the heme group (for the holo form) exhibited significantly lower fragmentation than the apo-myoglobin state. Both holo- and apo-myoglobin exhibited low fragmentation yields for the AGH helical core (reflecting its highest stability).

Table of Contents

Chapter 1: Probing the Unfolding of Myoglobin and Domain C of PARP-1 with Covalent Labeling and Top-Down 193 nm Ultraviolet Photodissociation Mass Spectrometry

1.1 Introduction

Biopolymer macromolecular structure determination remains a significant challenge owing to the large number of factors that influence molecular interactions, conformation, and folding. The structural characterization of proteins and their interactions with other proteins, DNA, RNA, and other ligands provides critical support in the arena of drug development as well as mechanistic insight into structure/function relationships. X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopic methods have dominated the field of structural biology due to their exceptional resolution. However, rather large amounts of proteins are required for both of these methods, and in particular for the X-ray methods numerous proteins do not crystallize or crystallize in ways that may be artifactual.[1,2] NMR measurements reflect the dynamic state of a protein in equilibrium and thus are especially versatile, although the molecular size limit (approximately 40 kDa) has restricted the range of proteins studied to date.[1]

These technological issues associated with NMR and X-ray methods have motivated the development and application of mass spectrometric strategies for characterization of protein structures and macromolecular complexes, thus providing access to a greater range of proteins. In particular, the scope of in-solution labeling techniques coupled with mass spectrometric detection has accelerated in the quest to discern low resolution protein structure information based on correlating mass shifts of fragment ions with the location of reactive or exchangeable sites of the backbone or specific amino acid side-chains.[3] Among the mass spectrometric methods, hydrogen deuterium exchange (HDX) gives the highest resolution information of the protein backbone. Covalent labeling, which typically targets specific amino acids via irreversible

reactions, is a popular complementary technique to HDX and targets side-chains rather than the protein backbone.[4,5] There are a myriad of covalent probes that have been employed successfully to study protein structure, function, and ligand binding.[6,7] The three most common covalent probe methods are based on oxidative,[8–11] carboxyl,[13] and primary amine [3,14–21] labelling reactions.[14,14–21] Some other less specific tagging techniques, such as diethylpyrocarbonate (DEPC) labeling and persistent carbene labeling, have proven useful for mapping protein topology as well.[22–24]

The strategy used in the present report is based on a primary amine labeling method. The most popular primary amine reaction employed for studies of surface topology is acetylation using acetic anhydride or a sulfo-NHS derivative of acetate.[18-21] Despite the low cost and small size of these acetylation reagents, removal of basic primary amine sites in proteins, may cause partial protein denaturation depending on the reaction stoichiometry. To circumvent this issue another amine modifier, S-methylacetimidate (SMTA), was developed which replaces the original basic sites with comparable basic groups, thus causing little to no detectable denaturation of the protein structure even at high levels of SMTA incorporation.[20,21,25,26] This reagent has been used to probe an array of proteins, including ribosomal complexes and most recently a viral capsid.[21,27–31] In the present study we report a similar primary amine derivatization reagent, S-ethylacetimidate (SETA) which maintains the same properties as SMTA but is easily synthesized.

Generally each of the chemical probe/mass spectrometry strategies mentioned above utilizes a bottom-up approach for characterization of the labeled proteins or protein complexes. This bottom-up approach entails proteolytic digestion of the covalently-labeled proteins followed by LC-MS/MS for separation, identification, and relative quantification of the resulting labeled and

unlabeled peptides in order to estimate the reactivities/accessibilities of particular sites. The bottom-up method has been a robust method for these studies; however; there are a few drawbacks.[32] The largest shortcoming evolves from the proteolytic digestion step of the workflow in which several proteases are typically employed to ensure complete coverage of the protein(s) via the peptide fragment maps and maximal detection and quantification of the labeled sites. Even the use of several proteases may be insufficient to produce full coverage of all labelled sites if they occur in peptides that are poorly ionized or elute in congested regions of the chromatographic profiles. Furthermore, protease activity may be impeded at labeled sites, creating gaps in sequence coverage. In an effort to alleviate the limitations of bottom-up methods, top-down strategies, which entail activation and dissociation of intact proteins, have gained significant traction in recent years with the increasing availability of high resolution/high accuracy mass spectrometers.[45] Because intact proteins are analyzed, both unequivocal measurement of the protein molecular weight (including the number of covalent labels) and the potential for full sequence coverage without critical gaps are possible. However, conventional fragmentation techniques such as CID do not always yield sufficiently high fragment coverage across the entire protein backbone, a limitation that mitigates one of the main advantages of the top-down approach for cases in which it is critical to pinpoint the sites of modifications or covalent labels. Alternate non-ergodic activation techniques, such as electron-capture dissociation (ECD) and electron transfer dissociation (ETD), have proven very effective for sequencing proteins as well as localizing PTMs, covalent labels, and sites of HDX,[33–35] although the performance of these electron-based methods varies considerably with protein charge state. An alternative new activation method for fragmentation of peptides and proteins is ultraviolet photodissociation (UVPD).[36,36–39] Recently, our lab has equipped an Orbitrap mass spectrometer

with a 193 nm excimer laser for UVPD of intact proteins.[39] It was found that the top-down UVPD fragmentation gave unprecedented sequence coverage of proteins up to 30 kDa with no charge state dependence.[40] This high energy activation technique is used in the presented study to afford the broad sequence coverage needed for mapping SETA incorporation in proteins for surface accessibility studies.

Two proteins were chosen for the present study, the very well characterized horse heart myoglobin and the C domain of the protein PARP-1 (PARP-C). Myoglobin, the first protein ever crystalized, is an oxygen-binding protein with a heme-group stabilized in the hydrophobic core with well solved structures by X-ray crystallography and NMR spectroscopy.[41,42] Mass spectrometry, analytical ultra-centrifugation and various spectroscopic studies have been used to examine the folding equilibrium of myoglobin as a function of varying buffer conditions such as pH and organic solvents.[12,43–45] The rich history of the structural biology of myoglobin makes it an ideal benchmark protein to validate SETA labeling in conjunction with top-down UVPD mass spectrometry in the present study.

The second protein, poly (ADP-ribose) polymerase 1 (PARP-1), is a major target of breast-cancer research with significant effort devoted to the development of more potent inhibitors.[46] This nuclear protein is responsible for reading DNA strands for damage; when damage is detected PARP-1 auto-modifies itself and other proteins via interaction with poly (ADP-ribose) (PAR) which signals, depending on the level of DNA damage, for apoptosis or DNA repair.[46] PARP-1 contains six domains, three of which are zinc finger domains known to interact with DNA. Removal of the third zinc binding domain (domain C) deactivates the PARP-1 protein, thus rendering it an essential domain for biological function.[47,48] Interestingly, domain C (PARP-C) does not bind to DNA on its own.[48,49] There are three solved structures of PARP-C

at this time, one from an NMR study and two based on crystal structures.[47–49] Of the two crystal structures, PARP-C crystallized by itself was solved as a dimer with an extended C-terminus,[49] whereas the second crystal structure was solved in complex with the first zinc binding domain, WGR, and catalytic domains with a close packed C-terminus,[47] which is in agreement with the results obtained from the NMR study. The disagreement between the NMR and crystal structures of the single PARP-C domain raises the question whether this discrepancy reflects a crystallization artifact.[48] In the present study, covalent labeling by SETA and solvent-mediated denaturation of PARP-C in conjunction with top-down UVPD-MS analysis is used for conformational characterization of PARP-C and to shed light on the past inconsistencies arising from analysis of the C-terminal region.

1.2 Experimental

1.2.1 Synthesis of S-ethylacetimidate:

SETA was synthesized by mixing 1 molar equivalent (m.e.) of dry acetonitrile with 1.4 m.e. of 4 M HCl in dioxane for 10 minutes and heating at 60 degrees to protonate the acetonitrile. Ethanethiol was added to a ratio of 1.2 (m.e.) and allowed to reflux overnight in a pressurized conical vial. The reaction was allowed to cool to room temperature and a small volume (~3 drops) of ethyl ether was added to ensure full precipitate formation. The white precipitate was filtered and washed three times with ethyl ether to remove impurities. This reaction yield was >80%. NMR and APCI mass spectrometric characterization of SETA are shown in **Figures 1.1** and **1.2**.

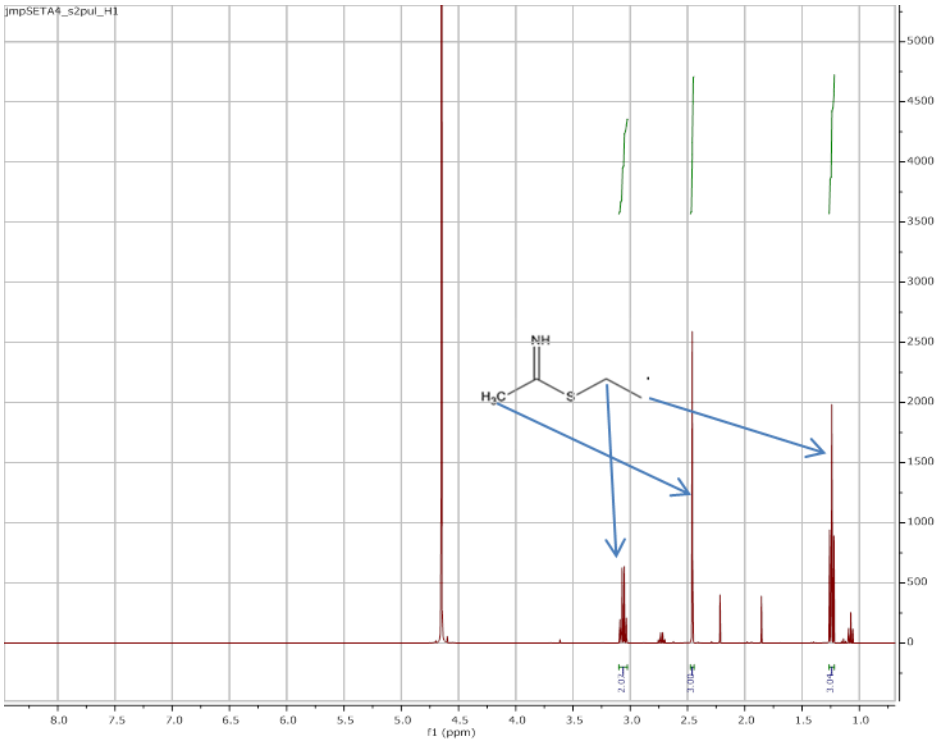


Figure 1.1 Resulting NMR from SETA synthesis. Corresponding hydrogen shifts are indicated with arrows.

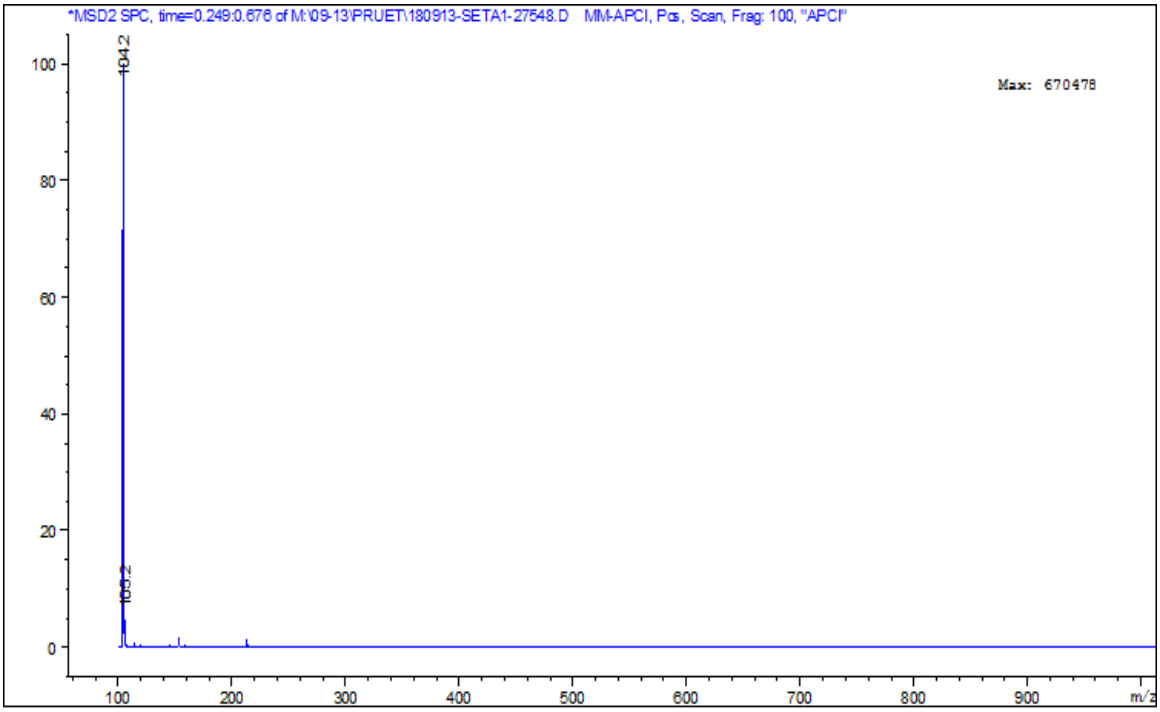


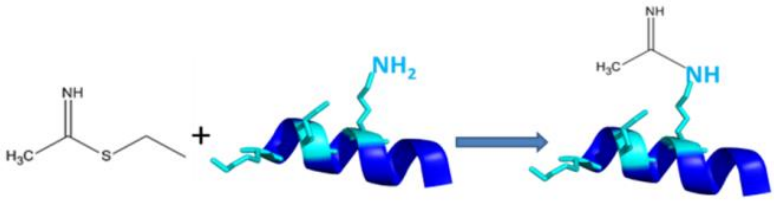
Figure 1.2 APCI mass spectrum of SETA. A mass of 104.2 Da was detected corresponding to the theoretical protonoated mass of SETA (104.195 Da).

1.2.2 Cloning, expression and purification of PARP-1 domain C:

Human PARP-1 domain C (PARP-C) was cloned and expressed according to similar procedures as described previously.[48] Briefly, DNA encoding residues 233–373 of human PARP-1 protein was cloned into a maltose binding protein (MBP) fusion vector (MalE-pET) containing a polyhistidine (His_{10}) tag and a TEV protease cleavage site. The resulting MBP fusion protein construct was overexpressed in *E. coli* BL21-CodonPlus (DE3)-RP strain and purified by Ni-NTA affinity chromatography. The N-terminal His_{10}-MBP was cleaved by 5% (w/w) His_6-tagged TEV protease during overnight dialysis and removed by a second Ni-NTA affinity column. The eluted PARP-C protein was then concentrated and dialyzed into 1X PBS buffer.

1.2.3 Protein Reactions:

60 µg of either PARP-C (16 kDa, containing 19 primary amines) or horse myoglobin (17 kDa, with 22 primary amines) (Sigma, St. Louis, Missouri) were diluted into each of three different buffer compositions containing 1X PBS buffer and acetonitrile. Compositions consisted of 0%, 25%, and 50% acetonitrile (ACN) with the remainder being composed of 1X PBS buffer at a pH of 7.4. The proteins were equilibrated for 15 minutes on ice. SETA stocks were prepared with each of the same respective buffers immediately before use. An equal volume of the SETA solution was added to each of the protein solutions to initiate the reaction at a ratio of 1:475 for PARP-C and 1:550 for myoglobin, with respect to protein:SETA molar ratio. All reactions were carried out for one hour on ice in triplicate. The reaction may be seen in Scheme 1.1. The reactions were quenched with an excess of 100x ammonium acetate relative to the concentration of SETA, vortexed, and then the proteins were separated from the quenched reagent using a 10 kDa molecular weight cut-off filter. The cleaned samples were stored in a -80°C freezer until analysis.



Scheme 1.1 Reaction of SETA with a deprotonated lysine.

1.2.4 Mass Spectrometry:

Each sample was diluted to 10 µM in 60:39.5:0.5, acetonitrile:water:formic acid prior to direct infusion into a Thermo Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a 193 nm excimer laser (Coherent Excistar XS). Each sample was infused at a rate of 1.2 ul/min with a spray voltage of +3.5 kV. ESI mass spectra were

collected at 120,000 resolution at m/z 400, averaged for 100 scans. MS/MS data was acquired by ETD and UVPD for the +19 charge state envelopes of PARP-C or myoglobin (including both the unmodified and all SETA-modified forms in the 19+ envelope). MS/MS spectra were averaged over 2000 scans at 120,000 resolution at m/z 400. The ETD activation time was 35 ms for myoglobin. The UVPD spectra were acquired at varying laser energies, from 1.5 mJ to 2.2 mJ, using a single pulse per scan. All UVPD experiments were carried out at 5 mTorr N_2 in the HCD cell.

1.2.5 Data Analysis:

Each ESI and UVPD mass spectrum was deconvoluted into neutral species with the Xtract algorithm provided by Thermo Fisher Scientific using a S/N threshold of 1.5 . The S/N threshold was set at a low value so that all low abundance modified fragment species produced by UVPD or ETD would be included in the data analysis. ESI mass spectra were evaluated manually to monitor the progress of the SETA labeling reactions. ETD MS/MS spectra were processed using ProSightPC 3.0. The deconvoluted UVPD spectra were processed by a modified version of ProSightPC 3.0 which allowed *a, a•, b, c, x, x•, y, y•,* and *z•* ions to be searched and identified. The fragment ion mass tolerance was set at 10 ppm for all processing. Both N- and C-termini fragment ions were searched for SETA modifications based on the addition of 41.02709 Da *n (n being the number of SETA modifications) to reveal fragment ions that contained the SETA label. Multiple iterations of this were undertaken upon variation of "n" until no additional ions were identified. Each of these iterations was exported to an Excel spreadsheet and was processed further organized by N or C-terminal ions.

Data analysis was undertaken in a manner similar to that reported by Pan *et al.*[33] Envelopes of fragment ions containing a particular unmodified diagnostic ion plus its modified

counterparts (i.e. containing one or more appended SETA moieties) were used to estimate SETA incorporation at each lysine residue within the protein. Weighted averages were calculated to estimate the incorporation of SETA in each identified fragment ion series. Only ions *a* and *a*• was considered for the N-terminus and *x, x*•*, y*, and *y*• for the C-terminus for the UVPD data. For the ETD spectra, only *c* and *z* ions were included. A pictorial representation of this data analysis is shown in **Figure 1.3** for the envelope of the a_{35} ion for unmodified and SETA-modified PARP-C. The red line expresses the weighted average (w_{av}) calculated for the modified ion series a_{35} relative to unmodified a_{35}. This weighted average conveys the average mass shift of the particular fragment ion (i.e. a_{35}) and thus directly reflects the number of SETA modifications incorporated in the collection of residues contained in each fragment ion (e.g. the a_{35} ion contains the first 35 amino acids of PARP-C, **Figure 1.3**). Weighted averages of each ion series were calculated as shown below:

$$W_{av} = \frac{\Sigma(product\ of\ \frac{m}{z}\ of\ each\ modified\ ion\ and\ its\ ion\ abundance)}{\Sigma(All\ ion\ abundances\ in\ modified\ ion\ series)}$$

The total SETA incorporation (S.I.) value for each modified fragment ion series was calculated by subtracting W_{av} from the monoisotopic theoretical mass of the corresponding unmodified fragment ion, then dividing by the mass of SETA (41.02709 Da).

$$S.I._{.total} = \frac{(W_{av} - mass_{theo.})}{mass_{SETA}}$$

This gives the total SETA incorporated for each fragment ion as SETA labelling is a cumulative process for the protein and therefore the fragment ions produced upon UVPD contain multiple primary amines (mostly lysine side-chains). The S.I. per individual lysine is calculated as the increase in SETA incorporation for each sequential lysine in the protein sequence, as follows:

$$S.I._N = (S.I._{total}\ \ at\ N\ lysine\ - S.I._k\ \ from\ N-1\ lysine)$$

where N is the number of the lysine in the sequence. For cases in which the fragment ion abundances are too low or there are gaps in sequence coverage that prevent calculation of individual SI values, then SI values are calculated for a pair or series of adjacent lysines where n is the number of lysines being averaged and N and M represent the positions of the lysines.

$$average\ S.I._{NM} = \frac{(average\ S.I.\ at\ M\ lysine\ - average\ S.I.\ from\ N\ lysine)}{n}$$

These values were tabulated, and the standard deviations were calculated based upon the replicates of the individual S.I. values per lysine. Variations in SETA incorporation values were evaluated using the Student's t-test to verify whether the variations as a function of buffer composition were significant.

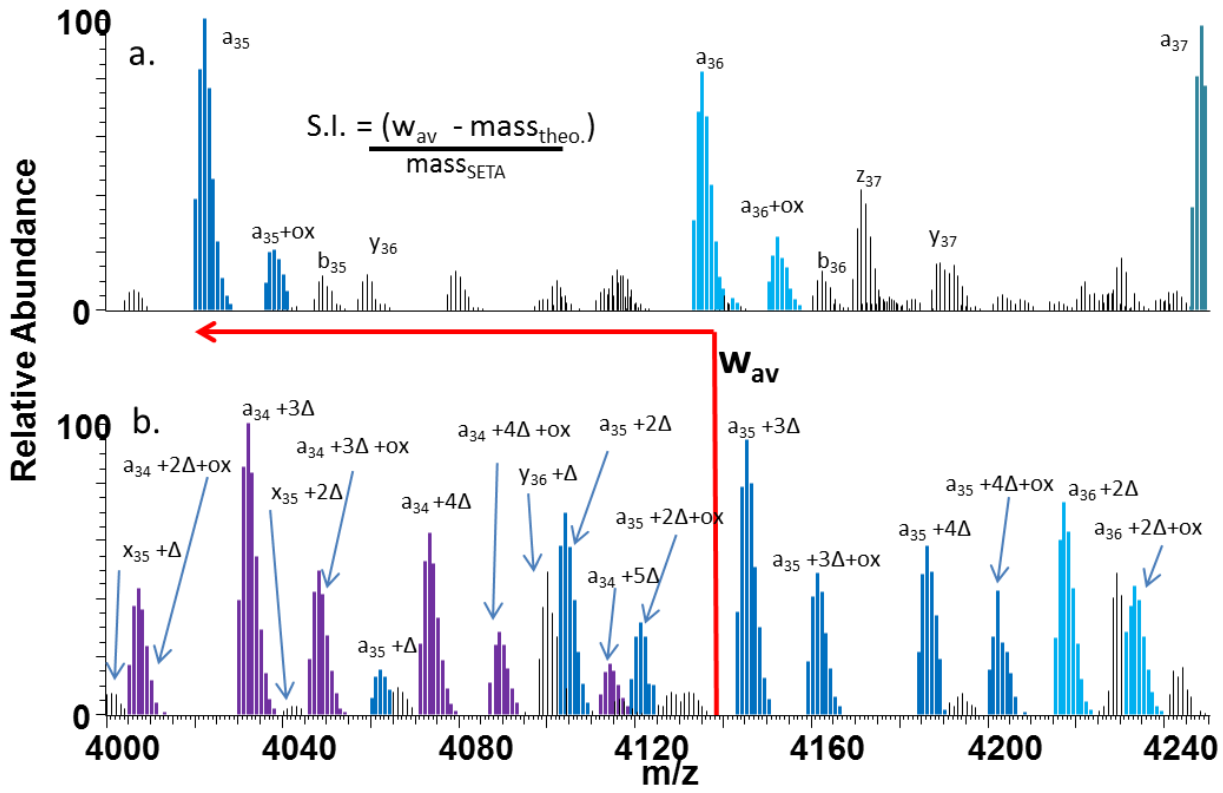


Figure 1.3 Decharged UVPD spectra of (a) unmodified and (b) SETA-modified Parp-C protein. The *a* ions have been color coded between panel a and b so that the modifications are more easily differentiated. An example of a weighted average (w_{av}) of SETA incorporation is highlighted with a red bar for the a_{35} ion series of SETA-modified Parp C relative to a_{35} of unmodified ParpC. Δ represents a single SETA modification of 41.02709 Da. ox represents the oxidation of a residue with a mass addition of 15.9949 Da.

1.2.6 Structural Analysis:

Two algorithms were used for predicting the solvent accessible surface area (SASA) and the local pKa values of each lysine: GetArea and PROPKA, respectively.[50–53] PROPKA also assigned strong hydrogen bonding interactions based upon their qualitifications. The parameters for GetArea were defined to include a water probe with a radius of 1.4 Å with no gradient calculations. Values from GetArea are in a percentage exposed calculated by the GetArea algorithm. Both NMR and crystal structures of myoglobin (1MYF and 1DWR) and PARP-C (2JVN and 2RIQ) were utilized as input structures for these algorithms. Prior to PROPKA

measurements the hydrogen atoms were removed from the NMR structures to make them more consistent with the crystal structures. Horse heart myoglobin lacks a solved NMR structure; in place of this the NMR structure of sperm whale myoglobin was used (1MYF). Both species of myoglobin (horse and sperm whale) structures are highly homologous with a backbone alignment of 0.831Å and sequence homology of 87.66%, allowing these structures to serve as suitable models in the present study.

1.3 Results/Discussion

For this study, UVPD was first used to characterize unmodified myoglobin and PARP-C to provide appropriate benchmark MS/MS results for each protein. For the unmodified proteins, UVPD yielded sequence coverages of 82% for myoglobin (19+ charge state) and 72% for PARP-C (19+ charge state), thus illustrating the broad and deep fragmentation afforded by UV photoactivation that facilitates pinpointing sites of modification. Each protein was subsequently reacted with SETA for 60 minutes in three buffer compositions ranging from 0% to 50% acetonitrile, followed by analysis by ESI-UVPD-MS. The same charge states were used for the subsequent UVPD-MS analysis of the SETA-modified proteins. The SETA incorporation values were determined based on the strategy described in the experimental section, and the results are summarized in **Figures 1.4** (PARP-C) and **Figure 1.5** (myoglobin) and **Table 1.1,** as well as **Table 1.2** (PARP-C) and **Table 1.3 and 1.4** (myoglobin) and supported with models (**Figures 1.6 – 1.9**), as described in the following sections.

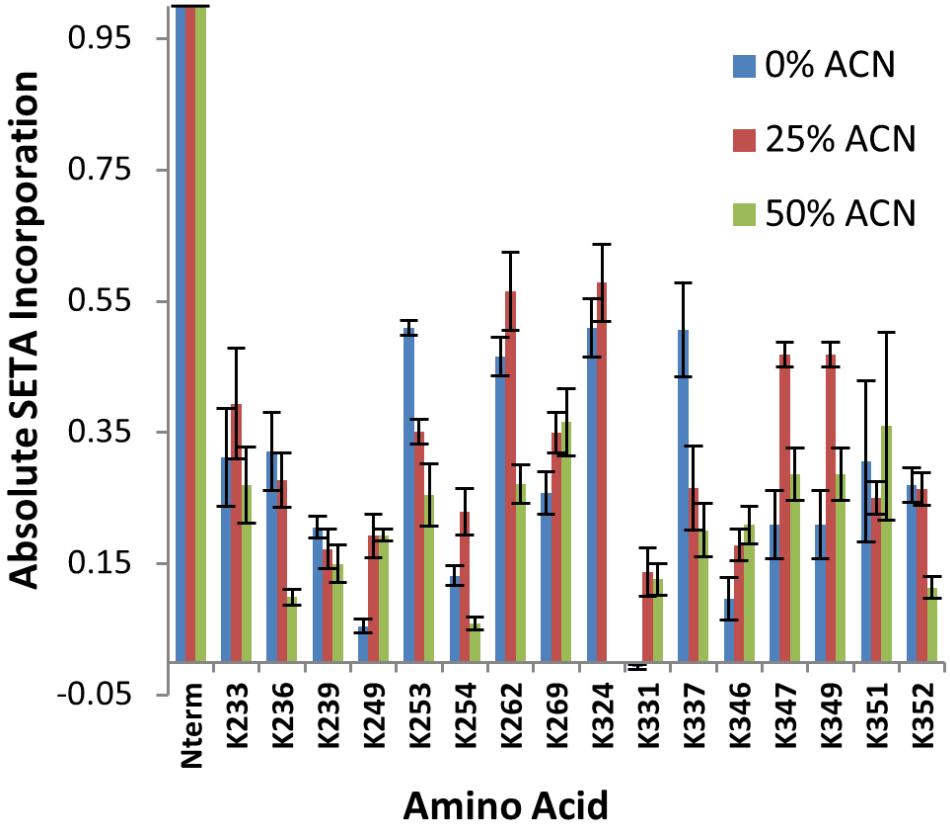


Figure 1.4 Absolute SETA incorporation values (n = 3) of PARP-C per lysine for the native (0% ACN), 25% ACN, and 50% ACN buffer conditions based on UVPD data. K324 was not quantified in the 50% ACN buffer due to lack of sequence coverage beyond K324. K305 and K320 were not observed. K347 and K349 are averaged together. K347 and K349 values are averaged together.

	0 %ACN	25 %ACN	50% ACN
Nterm	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
K233	0.31 ± 0.07	0.39 ± 0.08	0.27 ± 0.06
K236	0.32 ± 0.06	0.28 ± 0.04	0.10 ± 0.01
K239	0.21 ± 0.02	0.17 ± 0.03	0.15 ± 0.03
K249	0.05 ± 0.01	0.19 ± 0.03	0.19 ± 0.01
K253	0.51 ± 0.01	0.35 ± 0.02	0.25 ± 0.05
K254	0.13 ± 0.02	0.23 ± 0.04	0.06 ± 0.01
K262	0.47 ± 0.03	0.57 ± 0.06	0.27 ± 0.03
K269	0.26 ± 0.03	0.35 ± 0.03	0.37 ± 0.05
K324	0.51 ± 0.04	0.58 ± 0.06	NA
K331	-0.01 ± 0.00	0.14 ± 0.04	0.13 ± 0.02
K337	0.51 ± 0.07	0.27 ± 0.06	0.20 ± 0.04
K346	0.10 ± 0.03	0.18 ± 0.02	0.21 ± 0.03
K347	0.21 ± 0.05	0.47 ± 0.02	0.29 ± 0.04
K349	0.21 ± 0.05	0.47 ± 0.02	0.29 ± 0.04
K351	0.31 ± 0.12	0.25 ± 0.02	0.36 ± 0.14
K352	0.27 ± 0.03	0.26 ± 0.02	0.11 ± 0.02

Table 1.1 Absolute SETA incorporation values (n = 3) of Parp-c per lysine for the native (0% ACN), 25% ACN, and 50% ACN buffer conditions based on UVPD data. K324 was not quantified in the 50% ACN buffer. K305 and K320 were not observed. K347 and K349 are averaged together. K347 and K349 values are averaged together.

	Experimental		2JVN (NMR)			2RIQ (Crystal)		
Amino Acid	SETA Incorporation 100% aq	SETA incorporation trend	SASA	SHB location	pKa	SASA	SHB location	pKa
K233	0.31	No change	100	--	10.5	33	--	10.4
K236	0.32	Decreasing	81	--	10.3	94	--	10.4
K239	0.21	No change	86	--	10.5	77	D243	11.4
K249	0.05	Increasing	21	D281	10.7	32	D281	11.2
K253	0.51	Decreasing	65	--	10.5	65	--	10.1
K254	0.13	Variable	65	D250	10.7	56	E251	11.4
K262	0.47	Decreasing	50	--	10.4	46	--	10.4
K269	0.26	Increasing	93	Amide H K269, K305	10.4	88	--	10.3
K324	0.51	No change	74	--	10.7	53	--	10.5
K331	0.00	Increasing	65	Amide H R330, E332, K346	10.4	69	--	10.4
K337	0.51	Decreasing	70	E263	11.3	27	--	10.4
K346	0.10	Increasing	25	K331, E332 S343	9.7	44	D260	11.3
K347	0.21	Variable	88	--	10.5	52	--	10.1
K349	0.21	Variable	25	--	9.8	63	--	10.4
K351	0.31	No change	43	--	10.2	100	--	10.4
K352	0.27	Decreasing	100	--	10.4	96	--	10.4

Table 1.2 Summary of SETA incorporation trends (no change, increasing, decreasing, or variable as a function of acetonitrile composition) and predicted solvent accessible surface areas (SASA), strong hydrogen bonding interactors (SHB), and predicted pKa values for each lysine of PARP-C. The SASA values were calculated by GetArea. Local pKa values and SHB were calculated by PropKa. The ParpC structures were based on the solution NMR structure 2JVN and crystal structure 2RIQ. SHB locations were found by manual inspection in the PDB file 2JVN and 2RIQ. K347 and K349 are averaged values. K305 and K320 were not calculated.

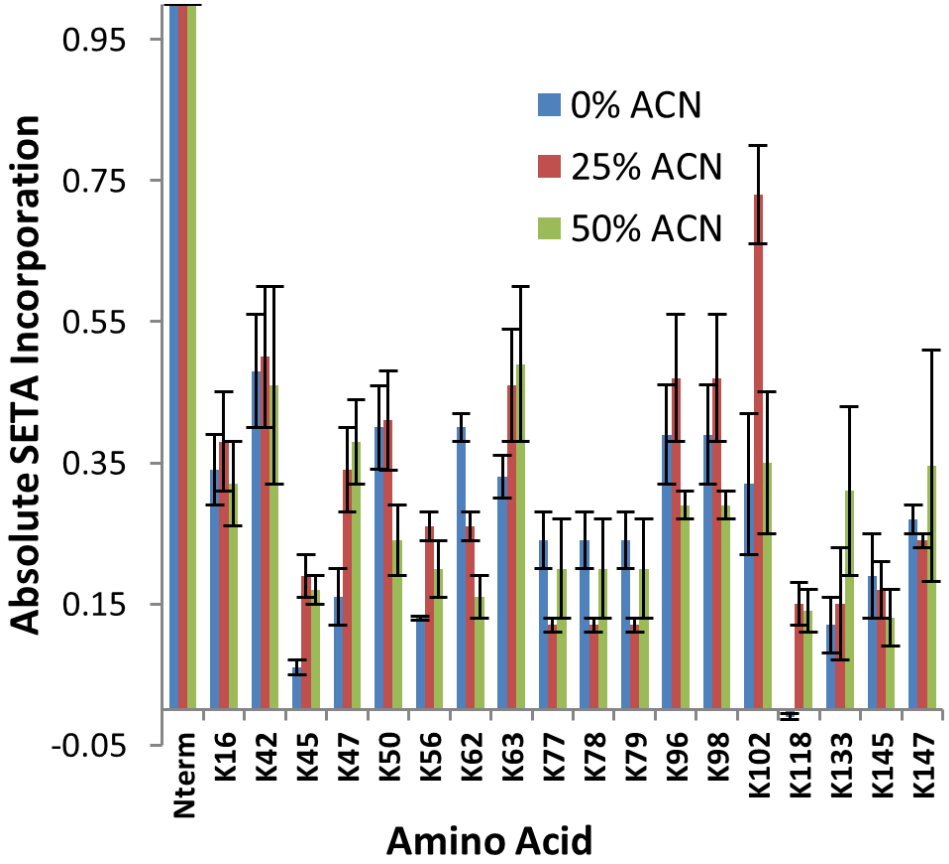


Figure 1.5 Absolute SETA incorporation values (n = 3) of myoglobin per lysine for the native (0% ACN), 25% ACN, and 50% ACN buffer conditions based on UVPD data. K86 was not observed. K77-K79 are averaged together. K96 and K98 are averaged together.

	0% ACN	25% ACN	50% ACN
Nterm	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
K16	0.34 ± 0.05	0.38 ± 0.07	0.32 ± 0.06
K42	0.48 ± 0.08	0.50 ± 0.10	0.46 ± 0.14
K45	0.06 ± 0.01	0.19 ± 0.03	0.17 ± 0.02
K47	0.16 ± 0.04	0.34 ± 0.06	0.38 ± 0.06
K50	0.40 ± 0.06	0.41 ± 0.07	0.24 ± 0.05
K56	0.13 ± 0.00	0.26 ± 0.02	0.20 ± 0.04
K62	0.40 ± 0.02	0.26 ± 0.02	0.16 ± 0.03
K63	0.33 ± 0.03	0.46 ± 0.08	0.49 ± 0.11
K77	0.24 ± 0.04	0.12 ± 0.01	0.20 ± 0.07
K78	0.24 ± 0.04	0.12 ± 0.01	0.20 ± 0.07
K79	0.24 ± 0.04	0.12 ± 0.01	0.20 ± 0.07
K96	0.39 ± 0.07	0.47 ± 0.09	0.29 ± 0.02
K98	0.39 ± 0.07	0.47 ± 0.09	0.29 ± 0.02
K102	0.32 ± 0.10	0.73 ± 0.07	0.35 ± 0.10
K118	-0.01 ± 0.00	0.15 ± 0.03	0.14 ± 0.03
K133	0.12 ± 0.04	0.15 ± 0.08	0.31 ± 0.12
K145	0.19 ± 0.06	0.17 ± 0.04	0.13 ± 0.04
K147	0.27 ± 0.02	0.24 ± 0.01	0.35 ± 0.16

Table 1.3 Absolute SETA incorporation values (n = 3) of myoglobin per lysine for the native (0% ACN), 25% ACN, and 50% ACN buffer conditions based on UVPD data. K86 was not observed. K77-K79 are averaged together. K96 and K98 are averaged together.

	Experimental		1MYF (NMR)			1DWR (crystal)		
Amino Acid	SETA incorporation 100% aq	SETA incorporation Trend	SASA	SHB location	pKa	SASA	SHB location	pKa
K16	0.34	No change	38	D122	10.1	22	D122	11.0
K42	0.48	No change	31	--	9.7	33	--	9.7
K45	0.60	Increasing	56*	D60,HEME	14.7*	54	D60	11.1
K47	0.16	Increasing	40	E41, D44	11.4	0.0	D44	11.1
K50	0.40	Decreasing	74	--	10.6	83	--	10.4
K56	0.13	Increasing	66	E54, D27, E59	10.8	58	--	10.2
K62	0.40	Decreasing	49	--	10.4	48	--	10.7
K63	0.33	Increasing	61	D60, E59	11.5	75	--	10.5
K77	0.24	Decreasing	82	--	10.4	51	E18	11.1
K78	0.24	Decreasing	40	--	10.3	46	E85	11.4
K79	0.24	Decreasing	49	--	10.1	54	--	10.5
K96	0.39	Decreasing	71	--	10.5	84	--	10.5
K98	0.39	Decreasing	88	--	10.3	65	--	10.2
K102	0.32	Variable	59	E105	10.4	64	E105	11.1
K118	0.00	Increasing	28*	D20,E27	13.6*	24	E27	11.4
K133	0.12	Increasing	43	E6, N-term	10.5	45	E6	11.1
K145	0.19	No change	16	D141	11.1	16	E83	11.6
K147	0.27	No change	66	--	10.2	97	--	10.5

Table 1.4 Summary of SETA incorporation trends (no change, increasing, decreasing, or variable as a function of acetonitrile composition) and predicted solvent accessible surface areas (SASA), strong hydrogen bonding interactors (SHB), and predicted pKa values for each lysine of myoglobin. The SASA values were calculated by GetArea. SHB locations and pKa values were calculated by PropKa. The myoglobin structures were based on NMR structure 1MYF and crystal structure 1DWR. PDB file 1MYF is of sperm-whale myoglobin with RMSD structure alignment of 0.831Å and a sequence homology of 87.66%. Entries marked with asterisks are based on arginine residues, not lysines, from the NMR structure. SHB locations were found by manual inspection in the PDB file 1MYF and 1DWR. K77,K78, and K79 are averaged values, as well as K96 and K98. K86 was not quantified.

1.3.1 Comparison of ETD and UVPD Activation:

The most abundant charge state (19+) of SETA-modified myoglobin (in 0% ACN) was analyzed by ETD as well as UVPD in order to evaluate the performance of UVPD relative to ETD for tracking the SETA modifications. The SETA incorporation results determined from the ETD and UVPD data are shown graphically in **Figure1.10** and tabulated in **Table 1.5**. Two features are notable in this comparison. First, the sequence coverage (in terms of the number of diagnostic sequence ions arising from backbone cleavages) is greater for UVPD than ETD, as evidenced by the lack of ETD-based SETA values for lysine residues K62 to K118 in **Figure 1.10**. In short, SETA incorporation values could not be calculated due to a lack of appropriate fragment ions in the ETD spectra (i.e. gap in sequence coverage). This finding has also been noted in comparative UVPD/ETD studies of other proteins,[40] and recapitulates the often broader sequence coverage obtained from UV photoactivation. Second, the SETA incorporation values for each residue are comparable for both the UVPD and ETD data sets (with one exception), confirming the utility of UVPD for detection and quantification of covalent labels relative to the well-established ETD method. The sole discrepancy between the UVPD and ETD results occurs for K47. From the ETD results, there was only a single fragment ion series covering the K47 to L49 residues, thus providing inadequate information for determination of the SETA incorporation value of K47. In contrast, complete coverage between K47 and L49 was obtained by UVPD, allowing more confident assessment of the SETA value for K47. These features of UVPD-MS make it a compelling strategy for pinpointing and quantifying chemical probe modifications of proteins in a top-down format.

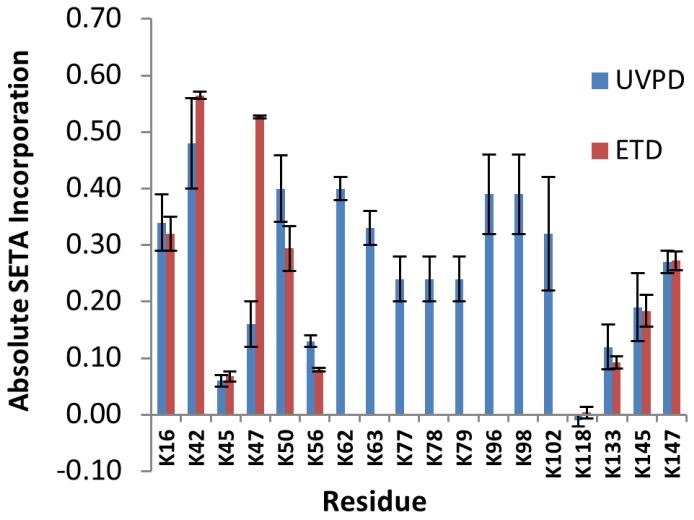


Figure 1.10 Comparison of SETA incorporation values (n = 3) determined by UVPD and ETD fragment ion results for the 19+ species of myoglobin labeled in the native state.

	UVPD	ETD
Nterm	1.00 ± 0.00	1.00 ± 0.00
K16	0.34 ± 0.05	0.32 ± 0.03
K42	0.48 ± 0.08	0.56 ± 0.01
K45	0.06 ± 0.01	0.07 ± 0.01
K47	0.16 ± 0.04	0.53 ± 0.00
K50	0.40 ± 0.06	0.29 ± 0.04
K56	0.13 ± 0.00	0.19 ± 0.01
K62	0.40 ± 0.02	----
K63	0.33 ± 0.03	----
K77	0.24 ± 0.04	----
K78	0.24 ± 0.04	----
K79	0.24 ± 0.04	----
K96	0.39 ± 0.07	----
K98	0.39 ± 0.07	----
K102	0.32 ± 0.10	----
K118	-0.01 ± 0.00	0.00 ± 0.00
K133	0.12 ± 0.04	0.09 ± 0.01
K145	0.19 ± 0.06	0.18 ± 0.03
K147	0.27 ± 0.02	0.27 ± 0.02

Table 1.5 Comparison of SETA incorporation values (n=3) detected by UVPD and ETD of the 19+ species of myoglobin.

1.3.2 SETA Incorporation as a function of increasing acetonitrile concentration:

The SETA incorporation values for each buffer condition as monitored by UVPD-MS are summarized in **Figure 1.5** (for myoglobin) and **Figure 1.4** (for PARP-C). Each protein exhibited complete N-terminal reaction with the SETA reagent (i.e. 100% SETA incorporation value), an outcome attributed to the pH of the solution (7.4) which is similar to the predicted pKa of the N-terminal amine of myoglobin (7.52) and PARP-C (7.72) from PropKa calculations. Undertaking the reactions at pH 7.4 means that the N-terminal amines are anticipated to be nearly completely deprotonated and thus have high reaction efficiencies with SETA. The pKa values of the primary amine side-chains of all the lysine residues are in the range of ~10 to ~12, so the majority are not expected to be deprotonated, and thus their reaction efficiencies should be more greatly influenced by accessibility. For the primary amine lysine side-chains, the SETA incorporation values ranged from near 0 (i.e. no or very low reactivity) to around 0.5 (high reactivity). Four lysines were found to have SETA incorporation values below 0.10 in the 100% aqueous buffer, specifically K45 and K118 for myoglobin and K249 and K331 for PARP-C, which categorizes them as the least reactive among the two proteins. Seven lysines were exceptionally reactive and displayed SETA incorporation values of 0.4 or higher in the 100% aqueous buffer: K42, K50 and K62 for myoglobin and K253, K262, K324, and K337 for PARP-C.

There are four general types of reactivity trends observed as a function of buffer composition (see **Figure 1.4** and **Figure 1.5**)**:** an increase in SETA incorporation; no change in SETA incorporation; a decrease in SETA incorporation; and variable reactivity (meaning an increase in reactivity in 25% ACN and a subsequent decrease in reactivity in 50% ACN). These specific reactivity trends for each residue are color coded and plotted onto the crystal structure

(**Figure 1.6**) or NMR structure (**Figure 1.7**) of myoglobin and PARP-C, respectively. Amines that have a decreasing SETA incorporation trend become less reactive upon exposure to the mildly denaturing conditions of acetonitrile. These sites include K50, K62, K96, and K98 of myoglobin and K236, K253, K262, K337, and K352 of PARP-C. Those lysines with increasing SETA reactivity include K45, K47, K56, K63, K118, and K133 for myoglobin and K249, K269, K331, and K346 for PARP-C. There were several amines which showed no significant change in SETA incorporation as a function of the acetonitrile composition of the buffer, including K16, K42, K77, K78, K79, K145, and K147 for myoglobin and the K233, K239, K324, and K351 for PARP-C. Finally, only four lysines showed variable behavior: K102 for myoglobin and K254, K347, K349 in PARP-C.

To rationalize the SETA incorporation trends of the lysines for each protein, two features of the chemical probe methodology must be understood: the factors that influence the reaction of the lysines with the SETA probe and the effects of acetonitrile denaturation on the tertiary structure of the proteins. In the context of the latter factor, acetonitrile has been shown to be a mild-denaturant at lower concentrations, disrupting hydrophobic interactions and causing loss of tertiary structure while maintaining secondary structural features, such as hydrogen bonding, until high concentrations of acetonitrile have been reached.[43,54–56] For some proteins, alpha helical content has been shown to be retained even in solutions containing more than 50% ACN.[43,54–56] In fact, myoglobin is one of the proteins whose structure has been well studied as a function of solvent composition. For example, upon exposure to 25% ACN, the non-covalent heme group of myoglobin, which is held in a hydrophobic pocket, is ejected due to the known unfolding of the loop region.[41–44]

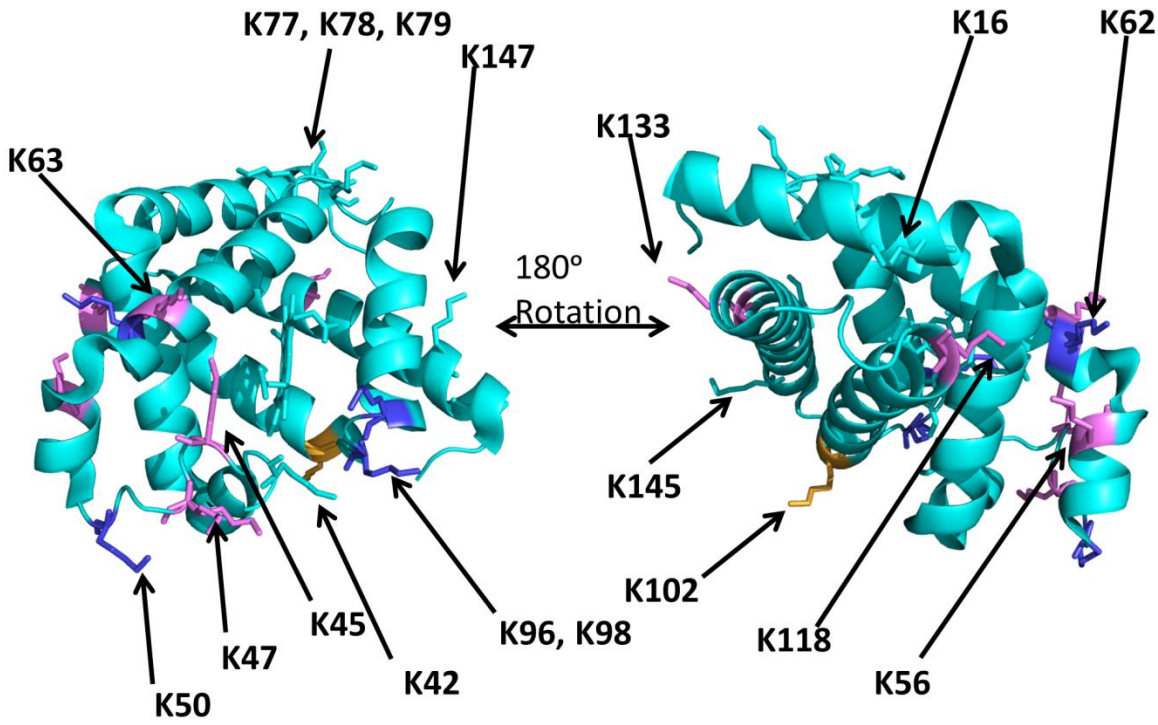


Figure 1.6. Myoglobin with color coded lysine residues which represent their SETA incorporation trend as a function of acetonitrile buffer composition with two views rotated 180°. Pink = increasing SETA incorporation, Blue = decreasing SETA incorporation, Cyan = no change in SETA incorporation, Orange = variable SETA incorporation. Crystal structure PDB ID used 1DWR. Hydrogen atoms were added by PyMol.

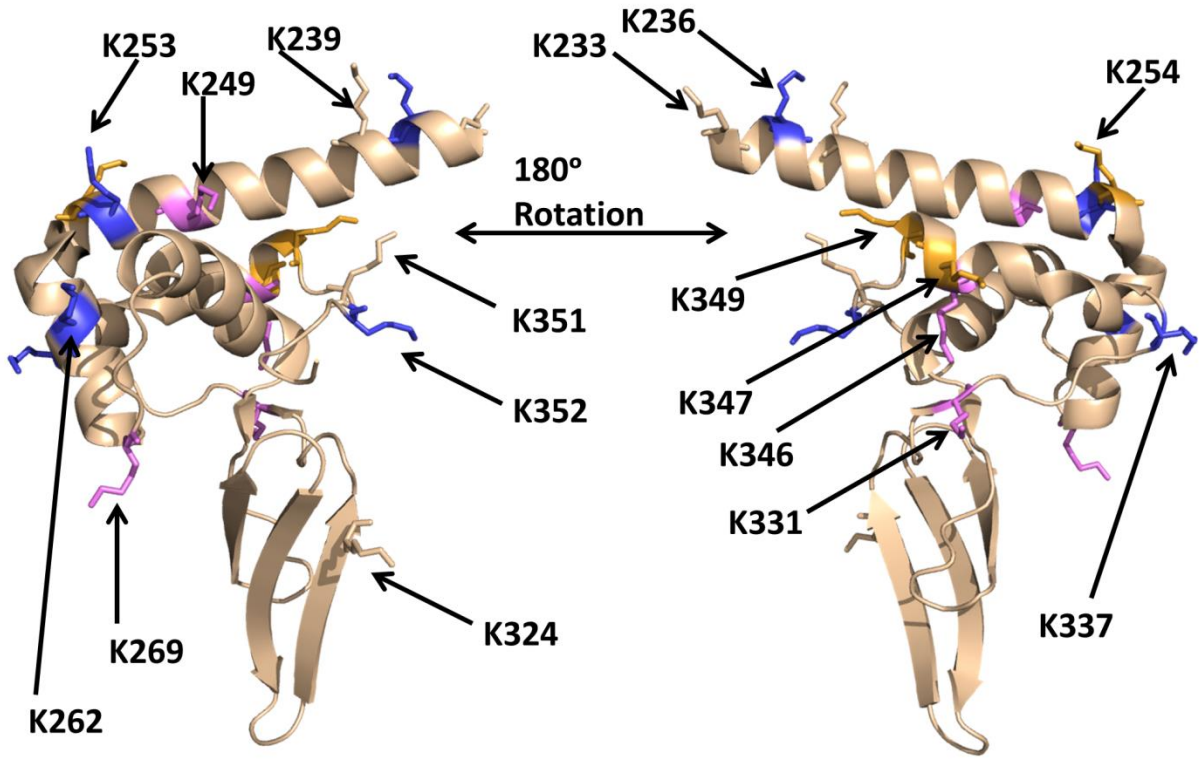


Figure 1.7 PARP-C with color coded lysine residues which represent their SETA incorporation trend as a function of acetonitrile buffer composition with two views rotated 180°. Pink = increasing SETA incorporation, Blue = decreasing SETA incorporation, Light Brown = no change in SETA incorporation, Orange = variable SETA incorporation. NMR structure PDB ID 2JVN, state 1 shown.

With respect to the intrinsic reactivity of a primary amine with the SETA chemical probe, two factors are expected to be most influential: the relative accessibility of the amine to the chemical probe which impacts its reaction probability and the local pKa of each lysine amine which determines its nucleophilicity. Residues with higher pKa values are less acidic and therefore are expected to more significantly favor the protonated (unreactive) form in solution. The local pKa is influenced by the interactions of the amine with other hydrogen-donating or hydrogen-accepting functional groups (i.e. polar contacts) and formation of salt bridges (i.e. electrostatics) which modulate the acidity of each primary amine. The pKa and the solvent accessible surface area (SASA) of each lysine side-chain are calculated and reported in **Table 1.2** and **Table 1.4** for PARP-C and myoglobin, respectively. In addition to the pKa and SASA values, noteworthy

hydrogen bonding interactions involving each lysine residue are included in **Table 1.2 and Table 1.4**. For this aspect, the NMR and crystal structures of each protein were inspected to identify the potential tertiary contacts between each lysine and interacting side chains of other residues (within a distance constraint of 4 Å from the gamma nitrogen of each lysine to the side-chain). The SETA incorporation values for the native proteins (in 100% aqueous buffer) and the trend in SETA incorporation as a function of increasing acetonitrile are also included in these tables.

When evaluating the SETA incorporation values for the various lysine amines of myoglobin, the trends observed as a function of acetonitrile composition of the solution can be rationalized by considering the two factors: local pKa and the SASA. Those amines possessing higher pKa values and low to modest SASA values, for example K249 from PARP-C and K45 from myoglobin, generally show an upward trend in SETA incorporation with increasing acetonitrile concentration. This result implies that these lysines engage in hydrogen bonds or salt bridge interactions with other residues of the protein in aqueous solution. As the acetonitrile content of the solution increase, local tertiary contacts are broken, in turn breaking the salt bridging or hydrogen bonding interactions which both decreases the pKa and increases the accessibility of these lysines. Conversely lysines that exhibit decreasing SETA incorporation as the acetonitrile composition increases are not predicted to be involved in hydrogen bonding and typically do not have elevated local pKa values, for example K50 from myoglobin and K352 from the PARP-C protein.

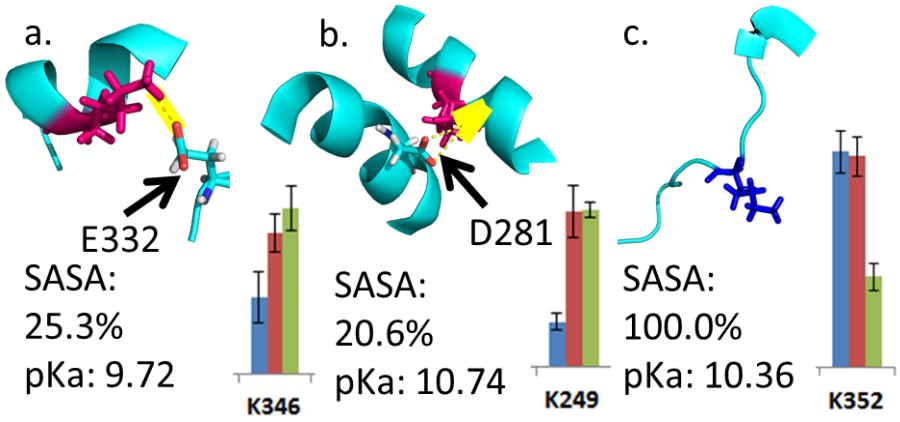


Figure 1.8 Specific lysine interactions are color coded (Pink = increasing absolute SETA incorporation, Blue = decreasing absolute SETA incorporation) to reflect their behavior as the acetonitrile content of the solution increases, with examples PARP-C (a,b,c). SASA and pKa values are also shown for the lysines, as well as the SETA incorporation values in bar graph form. The yellow lines represent hydrogen bonding interactions between the specific lysines and other side chains. Interactions were noted when the gamma nitrogen of lysine was positioned within 4.0 Å or less of the oxygen atom of the interacting side chain. Two lysines with increasing trends are shown (a,b) and exhibit positive interactions whereas one lysine (c) display no interactions . The interactions are specifically: (a) K346 to E332 (State 4/10), (b) K249 to D281 (State 7/10), and (c) K352 none (State 1/10). PDB structure 2JVN was used for PARP-C NMR structure. NMR structure states have been noted.

To illustrate examples of the different SETA incorporation behaviors in relation to local pKa values, SASA parameters and hydrogen-bonding interactions, three specific lysine interactions are showcased each for myoglobin (K45, K118, and K50) and PARP-C (K346, K249, and K352) in **Figure 1.9** and **Figure 1.8,** respectively. K45 of myoglobin has a mid-range SASA value (meaning mid-range solvent accessibility) of 54.9% based on the crystal structure and one of the higher pKa values (11.14) arising from its engagement in specific interactions with D60 and the heme group (**Figure 1.9a**). This residue displays a low SETA incorporation value (0.06) in the 100% aqueous buffer, a value that is consistent with an unreactive residue. Upon partial denaturation upon addition of acetonitrile, the local hydrophobic interactions are disrupted which decreases the pKa of K45, ultimately contributing to the increase in its SETA reactivity. This is a prime example of how the combination of the high pKa value and

involvement in tertiary interactions suppresses the reactivity of the lysine amine in its native state. K118 exhibits no reactivity with SETA in 100% aqueous buffer, a result that reflects both its low SASA value of 23.5% and its high pKa of 11.44 (**Figure 1.9b**). As the acetonitrile content of the buffer increases, the SETA incorporation value of K118 increases significantly, implying that the exposure of this residues increases due to partial unfolding. Based on its large SASA value of 83.4% and lower pKa value (10.43), K50 of myoglobin represents one of the more exposed and more acidic lysine residues (**Figure 1.9c**). K50 has a relatively high SETA incorporation value (0.40) in the 100% aqueous buffer, but the reactivity decreases by nearly a factor of two in 50% acetonitrile, indicating a significant change in its local environment that is reflected by its decreased SETA reactivity.

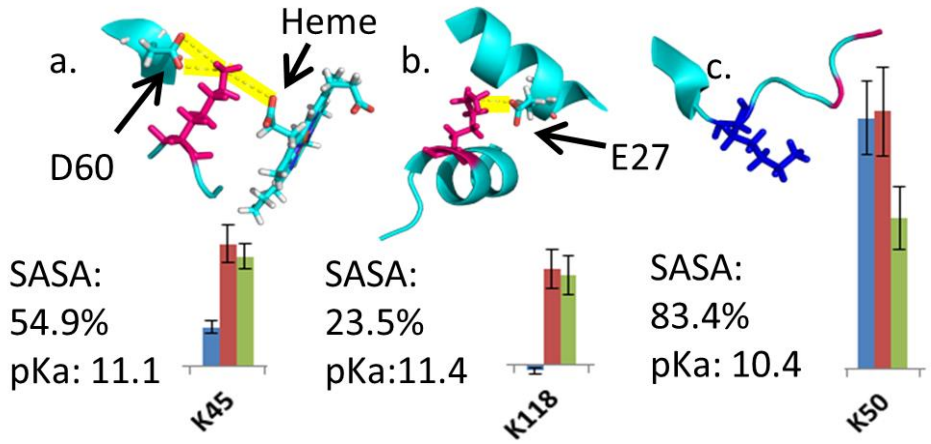


Figure 1.9 Specific lysine interactions from myoglobin (a,b,c) are color coded (Pink = increasing absolute SETA incorporation, Blue = decreasing absolute SETA incorporation) to reflect their behavior as the acetonitrile content of the solution increases. SASA and pKa values are also shown for the lysines, as well as the SETA incorporation values in bar graph form, where blue is 0% ACN, red is 25% ACN, and green is 50% ACN. The yellow lines represent hydrogen bonding interactions between the specific lysines and other side chains. Interactions were noted when the gamma nitrogen of lysine was positioned within 4.0 Å or less of the oxygen atom of the interacting side chain. Two lysines with increasing trends are shown (a,b) and exhibit positive interactions whereas one lysine (c) display no interactions . The interactions are specifically: (a) K45 to D60 and Heme, (b) K118 to E27, (c) K50 none. The PDB structures used were 1DWR for myoglobin crystal structure.

For PARP-C, K346 is a particularly interesting case. This residue has both a low SASA value (25.3%, low solvent accessibility) and a low pKa value (9.72, more acidic than other lysines) (**Figure 1.8a**). K346 has a very low SETA incorporation value in 100% aqueous buffer, and its reactivity increases but remains relatively low as the acetonitrile content of the solution increase. This behavior suggests that even with a low pKa, which should enhance the reactivity of the residue, the low solvent accessibility prevents adequate access to the SETA chemical probe. The lack of reactivity of K346 even in 50% ACN suggests that SETA accessibility remains low as though little structural reorganization occurs. K249 has a relatively high pKa (10.74) and a low SASA (20.6%) which results in low reactivity in the native 100% aqueous buffer composition. Upon mild denaturation by acetonitrile, the reactivity of K249 increases due to the disruption of the tertiary contacts between K249 and the two helical regions that buttress K249. The behavior of K352 of PARP-C mirrors that observed for K50 of myoglobin. In each case these two lysines are predicted to have very high solvent accessibilities in aqueous solution and exhibit a marked decrease in reactivity in 50% acetonitrile, thus suggesting a substantial change in their local environments. The fact that both "buried" (e.g. K45 and K118 for myoglobin and K249 and K331 for PARP-C) and "exposed" (e.g. K63 for myoglobin and K269 for PARP-C) residues may show increases in SETA incorporation as a function of acetonitrile buffer composition suggests that these residues may share orientations or local topological features that are responsive to denaturation.

Two of the variable behavior lysines have an interesting structural element in common in that they interact with acidic residues that exist on the same alpha helix as the lysine. K102 (myoglobin) may interact with E105 which is located on the same alpha helical element. These residues are highlighted in **Figure 1.11**. K254 of PARP-C may interact with D250 of the alpha

helix as seen in the NMR structure or with E251 in the crystal structure (**Figure 1.12**). The SETA reactivity of each of these residues increases in the 25% ACN buffer composition, implying that these lysine-helix interactions may be disrupted and thus allow higher reactivity with the SETA probe. The amines may make other non-native polar contacts in solution upon denaturation which may explain the variable SETA incorporation behavior seen for those two residues.

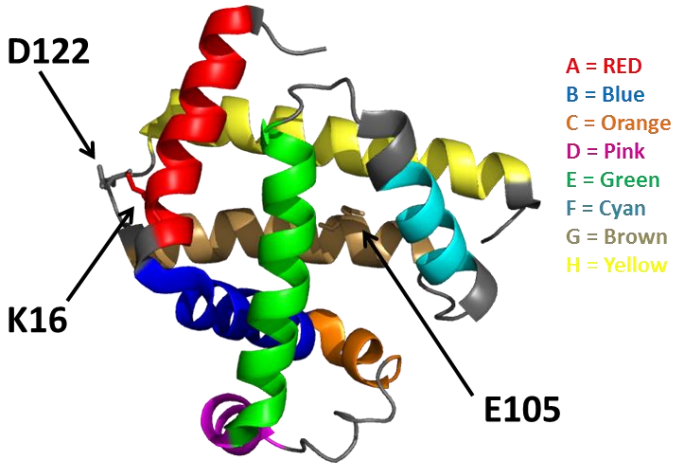


Figure 1.11 Crystal structure of horse heart myoglobin with color coded helices with respect to the legend above. Gray is for loop regions between the helices. These helixes have been defined as in Johnson et al. (1994). Specific residues have been pointed out and were discussed in the manuscript. PDB structure: 1DWR

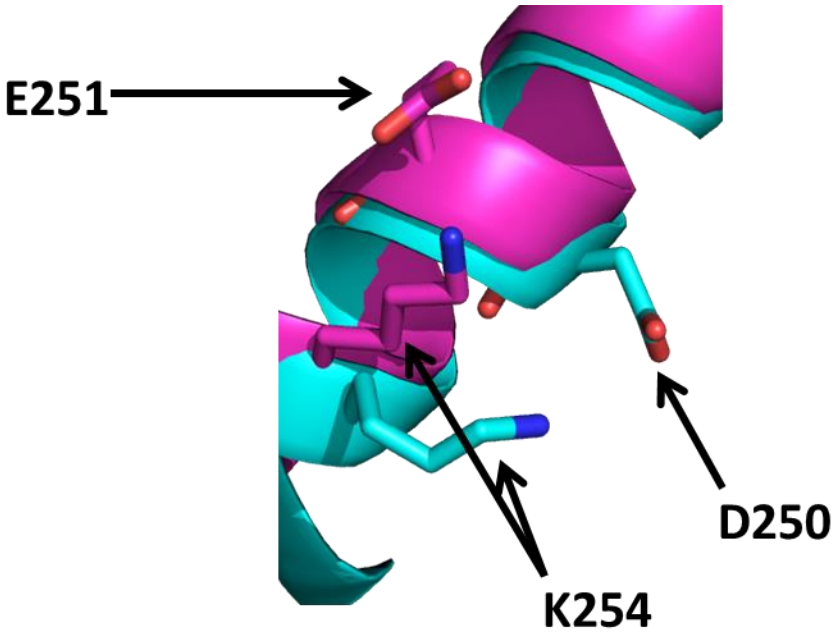


Figure 1.12 Zoom in of K254 showing interactions with D250 in the NMR structure (2JVN) (cyan colored) and E251 in the crystal structure (2RIQ) (purple colored). The NMR structure is in state 9. Interactions were found below a 4.0 Å cutoff.

1.3.3 Myoglobin Unfolding:

For the model system of horse heart myoglobin, previous studies have reported that increasing the acetonitrile composition causes the loop region to unfold and disrupts the hydrophobic pocket of myoglobin, thus resulting in the ejection of the heme group.[12,43–45] The unfolding mechanism deciphered upon acidic denaturation based on NMR spectroscopic studies and upon acetonitrile denaturation (27% ACN at pH 9.3) by HDX mass spectrometry is consistent with the observations in the present study. In brief, the reported mechanism shows that the helices A, G, H and partial helix B fold first into the core or scaffolding for the duration of the folding process, remaining folded in the molten globule state. This is followed by folding of the C, D, and E helices which comprise the apo-myoglobin native state. With the subsequent addition of the heme group, the F helix folds.[42,44] These helices can be found demarcated in **Figure 1.11**. In the present study K16, located on helix A, shows no change in SETA reactivity

as a function of acetonitrile denaturation but engages in a strong hydrogen bond to D122 located on the loop between helices G and H. The absence of change in SETA reactivity of K16 is consistent with the three helix A-G-H core remaining folded during mild denaturation. The secondary folding region (C,D,E) which contains the loop fold from helix C to D is likewise corroborated with the SETA reactivity data for residues K45, K47, K50, and K56. Each residue, with the exception of K50, were found to have contacts with acidic residues within this region. The increasing SETA reactivities (with the exception of K50) upon acetonitrile denaturation echo the established model of folding. The decrease in SETA reactivity of K50 in 50% ACN suggests the formation of alternative non-native contacts within the molten globular state. The significant increase in SETA reactivity of K118 (G helix, which interacts with E27 on the B helix, see **Figure 1.9b**) suggests that the B helix unfolds as well during acetonitrile denaturation. The exceptionally stable core of A-G-H may explain the unusual behavior of K102 with is located on the outer edge of the G helix. In a very recent conformational study it was observed that the c-terminal region of the H helix which interacts with the region of the G helix containing K102 is partially unfolded during acid denaturation.[57] The significant increase, then decrease, in SETA reactivity of K102 may reflect the increase in accessibility of K102 in 25% ACN, from the H helix unfolding, followed by formation of new contacts with the C-terminus, H helix, or G helix in the molten globule state (in 50% ACN). The N-terminus, as well as E6, interacts with K133 located on the H helix. K133 shows an increase in SETA reactivity as the acetonitrile concentration increases, an outcome consistent with the unfolding of the N-terminus region upon denaturation and concomitant greater accessibility of K133. In general, the SETA reactivity trends are accordant with the well-established unfolding pathways of myoglobin, thus validating the SETA/UVPD-MS strategy for investigation of protein folding.

1.3.4 PARP-C Implications:

The SETA reactivity results of PARP-C, particularly for K331, K337, K346, K347 and K349, support the monomeric structure of the C-domain of PARP-C solved by NMR spectroscopy. The monomeric structure of the alpha helical region of the C-terminus (F339 to L348) elucidated by NMR spectroscopy was found to engage in hydrophobic interactions with the N-terminal alpha helix, thus resulting in a more compact structure.[48] K346 (C-terminal region), which interacts with E332 (loop), showed an increase in SETA reactivity upon acetonitrile denaturation. The tertiary polar contact between K346 and E332 could not exist in the alternative dimeric structure based on distance constraints. In the dimeric structure, K346 was proposed to interact with D260 instead of E332, which could lead to a similar change in SETA reactivity. K337 showed a significant decrease in SETA reactivity upon acetonitrile denaturation. An increase in reactivity upon denaturation would be expected for K337 in a dimeric form based upon its initial low predicted solvent accessibility value of 27%. Additionally K337 exhibits one of the greatest SETA incorporation values (0.51) in 100% aqueous buffer, which is in agreement with the prediction for the monomeric, not dimeric, structure. K331 was found to be completely unreactive with SETA in the 100% aqueous buffer conditions. In both the monomeric and dimeric structures the predicted SASA of K331 is around 65%, but in the dimeric structure this residue does not interact with any other residues and therefore should exhibit high reactivity. K331 engages in many contacts in the monomeric form, including ones with the amide hydrogen of R330, with E332, and a possible interaction with K346. The SETA reactivity of K331 is therefore consistent with the monomeric form. The two residues in the C-terminus region, K347 and K349, that collectively have variable SETA reactivity, are not

predicted to participate in any hydrogen bonds; however, the solvent accessibility of K349 is low (25%) and that of K347 is high (88%). The fact that these residues show an overall increase in SETA reactivity in 25% ACN may be indicate that the accessibility of K349 increases during denaturation. If this domain that contains K347 and K349 existed in the dimeric form as suggested by the crystal structure, these residues would both be highly exposed and should not show the observed variation in SETA reactivity. The subsequent decrease in SETA incorporation of K347 and K349 as the ACN content increases from 25% to 50% is attributed to the formation of molten globule-like state where the side chains have made non-native contacts either through hydrophobic interactions of the alpha helix or electrostatically. Cumulatively the SETA incorporation results suggest that PARP-C exists in a monomeric form in solution and that the dimeric form of this domain may thus be an artifactual result of the crystal packing process.

1.4 Conclusions:

High-mass accuracy top-down UVPD-MS in conjunction with covalent labeling with an amine-selective SETA probe was used to monitor the reactivity of lysine side-chains of two proteins, thus providing a means to evaluate conformational changes as each protein unfolded during denaturation. Owing to its more extensive sequence coverage, UVPD outperformed ETD for characterization of intact modified proteins in which each region of the protein where there are missed cleavages results in uncertainties in calculating the SETA reactivity of the lysine residues. The UVPD-MS method was validated by examination of myoglobin during mild denaturation, and the results were congruent with previous HDX and NMR studies that probed the folding mechanism of myoglobin. The SETA labeling/UVPD-MS results for PARP-C indicated that PARP-C exists in a monomeric form in solution. Furthermore, inspection of the SETA incorporation trends at varying degrees of denaturation revealed that a complex

relationship for lysine reactivity involving not only the solvent accessibility of each lysine but also its local pKa, a parameter influenced by the formation of polar contacts with other residues. This approach can be used to evaluate the impact of denaturation on the involvement of side chains in tertiary contacts or polar interactions.

Chapter 2: Structural Characterization of Holo and Apo-Myoglobin in the Gas Phase by Ultraviolet Photodissociation

2.1 Introduction

Protein and macromolecular structure characterization, including the elucidation of protein-protein and protein-ligand interactions, remain significant technological challenges and are critical linchpins in the field of structural biology. In this context, the two most successful techniques for determination of protein structures are X-ray crystallography and nuclear magnetic resonance (NMR) spectrometry.[1,2] While these techniques produce unparalleled information about the core structures of proteins, each has specific drawbacks that have propelled the development of other methodologies. For example, crystallization of proteins can be a tedious process that may not always produce a protein crystal even after evaluation of hundreds of conditions. If the proteins are successfully crystalized, the resulting crystal structures reflect static structures which are not necessarily identical to ones in cellular environments and may suffer from artifactual packing.[2] NMR spectrometry captures the dynamic structures of proteins but has a practical mass limit of approximately 40 kDa.[1]

Due the nature of these limitations, a number of mass spectrometric methods have been developed for the structural characterization of proteins with high sensitivity and low sample demands. Some of these methods entail using various chemical probe methods on proteins in solution, followed by mass spectrometric detection of products. Hydrogen deuterium exchange (HDX) of the amide hydrogens and covalent labeling of specific amino acids remain at the forefront of mass spectrometric-based strategies for structural biology applications.[58] In these methods, the proteins labelled in solution are typically analyzed by a bottom-up type proteomic

workflow which entails digestion of the proteins by one or more proteases, separation of the resulting peptides by liquid chromatography and identification by mass spectrometry. The abundances of the modified peptides, which presumably reflect the accessibility of the HDX reagent or chemical probe to specific peptide sites, are quantified relative to the unmodified counterparts to reveal information about the tertiary structure of the protein.[59] This process can be simplified by adoption of a top-down workflow which alleviates the proteolytic digestion step but requires a high-resolution/high mass accuracy mass spectrometer to map the incorporation of the covalent labels.[33,60] For the top down approach, the deuterium-labelled or probe modified proteins are typically transferred to the gas phase by electrospray ionization (ESI) and fragmented by collisional induced dissociation (CID) or its analog higher energy collisional dissociation (HCD), electron-based activation (ECD or ETD), or ultraviolet photodissociation (UVPD), then the modified sites are determined based on interpretation of the abundances of resulting fragment ions.[33,60] Denaturing ESI conditions are typically used to maximize sensitivity. While the top down/labelling approach offers some advantages to the bottom-up methods, the use of any type of labelling method can introduce artifacts during data analysis.

An alternative to labelling methods is to use native spray conditions to ionize proteins for direct analysis in the gas phase. Although the structures of proteins in the gas phase remain a matter of debate and such studies raise concerns about conformational integrity, investigation of intact proteins may offer new insight as well as correlations with known solution structures. It has been shown previously that tertiary intra-protein, protein-ligand and protein-protein interactions can be retained in the gas phase using gentle ESI conditions and buffered solutions.[61–64] Several groups have undertaken HDX on the presumed "native" proteins and protein complexes and provided convincing evidence that the gas-phase structures mirror the

known crystal or solution structures obtained by X-ray crystallography or NMR.[65,66] However, hydrophobic interactions are not well maintained in the gas-phase native structures, and thus electrostatic and van der Waals type interactions play a far more important role. Ion mobility studies have also shown that ions in lower charge states may exhibit similar radii of gyration to that of their solution counterparts.[57,67] These studies have also demonstrated that as the charge states increase, the proteins are more elongated, thus showing the impact of charge on denaturation in the gas phase.[68,69]

There have been a handful of recent studies in which unlabeled native-like proteins have been interrogated by electron-based activation methods, ETD and ECD, to obtain structural insight.[70–76] The conclusions from these MS/MS-based studies have been compelling. For example, it was discovered that strong salt bridges were not disrupted by ETD, thus enabling discrimination of different types of salt bridges in the gas phase.[71] It was found that specific enhanced backbone cleavage sites could be correlated with regions of the proteins that were known to have high B-factors (a measure of flexibility) from crystallographic studies, suggesting that these regions are also quite disordered (and more flexible) in the gas phase and thus more prone to cleavage.[72,73,77,78] Similarly, the abundances of the fragment ions produced by ECD or ETD have been related directly to the stabilities of specific regions of the proteins.[77,79–81] There has been one recent report of characterization of native protein complexes by UVPD, ETD and HCD for which UVPD revealed the greatest information about the primary sequences of the proteins in the complexes and also produced non-covalent fragment ions comprised of a portion of the protein bound to the ligand.[82] In general, 193 nm UVPD has proven to be an exceptional activation technique for obtaining very high sequence coverage of the protein backbone (for both native and denatured proteins).[40,60,83,84] HCD has been applied to top-down analysis of intact

proteins (denatured),[40,83,85] including one study that showcased the fragmentation of the transmembrane section of integral membrane proteins under denaturing conditions.[86]

The present study focuses on evaluating the application of 193 nm UVPD to characterize the gas-phase structures of native proteins, as demonstrated for horse heart myoglobin in both its holo and apo (heme-bound) states. Horse heart myoglobin was the first structure to be solved by X-ray crystallography and has become the most popular model protein for development and assessment of new structural methodologies both in apo and holo forms.[41,42,44,60,87,88] However, an apo crystal structure has yet to be published. Past studies of myoglobin have been undertaken using an array of mass spectrometric strategies, such as the use of solution phase HDX for examination of structure and kinetic unfolding and ion mobility measurements for investigation of protein cross-section, as well as many computational studies on the stability of the helices in myoglobin.[42,44,45,60,89–91] As with past ECD/ETD experiments in which fragment ion abundances were related to the stabilities of particular regions of proteins,[63,74,79,81] the same types of correlations are developed in the present study from the examination of UVPD results. The impact of the charge state of the proteins is also examined to determine whether charge-mediated unfolding/elongation is reflected in the resulting fragment ion distributions, and HCD results are evaluated relative to the UVPD results to allow a comparison of the trends observed for a more conventional collision-based activation method (HCD) to photodissociation.

2.2 Experimental:

2.2.1 Mass Spectrometry and Data Collection:

A 5 μM solution of holo-myoglobin (Sigma, St. Louis, MI) was prepared from lyophilized powder in 5 mM ammonium acetate at pH 5.5 prior native spray. Apo-myoglobin was prepared by butanone extraction as previously described from lyophilized holo-myoglobin.[92]

After heme extraction the protein was diluted to the same conditions as the holo-form to facilitate ionization of the ligand-free myoglobin in its native state. Denatured apo-myoglobin was generated by infusing 5 μM myoglobin in a solution of 59.9% acetonitrile, 40% water, and 0.1% formic acid at a rate of 1.5 ul/min. All MS/MS analysis was undertaken using a Thermo Scientific Orbitrap Elite mass spectrometer (Bremen, Germany) outfitted with a 193 nm Coherent Excistar excimer laser for UVPD in the HCD cell as described previously[30] and a Prosolia electrospray ionization source (Indianapolis, IN). A voltage of 1.7 kV was applied for ESI, and the heated capillary temperature was maintained at 150°C. The range was monitored from m/z 220 to 4000 for both survey ESI-MS and all MS/MS scans. Each of the top three most abundant (8+, 9+, and 10+) charge states of the holo and apo protein, as well as the 20+ charge state of the denatured protein, were selected and activated by UVPD. Ions were isolated using a window of 10 m/z. 100 scans were averaged. All results were replicated three to four times. All MS/MS were collected at a resolution of 480K resolution at m/z 400. UVPD was performed using a single 1.5-2.0 mJ laser pulse in the HCD cell at a nitrogen pressure of approximately 10 mTorr. HCD was performed between 30 to 35% NCE with aforementioned settings.

2.2.2 Data Analysis:

The MS/MS spectra were deconvoluted using the Xtract algorithm with a S/N ratio of 2 and fragment ion identification tolerance of 10 ppm. The deconvoluted spectra were processed using a custom version of ProsightPC 3.0 which searched each spectrum using the nine ion types generated by UVPD (*a, a+1, b, c, x, x+1, y,Y, z*). The identified ion list was tabulated in an Excel format, and the normalized fragmentation yield per amino acid backbone cleavage site was calculated as follows:

$$\textbf{Normalized Fragment Yield} = \frac{\textbf{Σ Abundances of ions from specific amino acid}}{\textbf{Σ Abundances of all identified ions from Prosight}}$$

In this approach, all product ions arising from backbone cleavages that occur N-terminal to a specific amino acid (yielding a_n, b_n, and c_n ions) are summed with all the complementary C-terminal product ions arising from the same backbone cleavage site (yielding complementary x_{R-n} , y_{R-n}, and z_{R-n} ions) where R is the total number of amino acids in the protein. For example, the abundances of the a_{70}, b_{70} and c_{70} ions are summed with the abundances of the x_{83}, y_{83}, and z_{83} ions, each representing cleavage between residues 70 and 71 for myoglobin (which contains 153 amino acids). In this way, cleavages are categorized via a specific backbone position that is assigned to an amino acid local to the cleavage site. The MS/MS spectra for holo-myoglobin were additionally searched with inclusion of a +615.1695 Da shift corresponding to heme ligand adduction. The identified heme-containing fragment ions were summed with the corresponding apo-fragments from the same spectrum. The resulting fragment yields were also averaged into well-defined demarcated regions and then divided by the length (in terms of number of constituent amino acids) of the regions. These regions are as follows N-terminus (residues 1-3), A-helix (residues 4-19), B-helix (residues 20-35), C-helix (residues 36-42), CD loop (residues 43-50), D-helix (residues 51-57), E-helix (residues 58-77), EF loop (residues 78-85), F-helix (residues 86 to 94), FG loop (residues 95 to 99), G-helix (residues 100-118), GH loop (residues 119-123), H-helix (residues 124-150), and C-terminus (residues 151-153).[41,42,93] Effectively this gives a "fragmentation density" per structural region, a categorization that is beneficial for facilitating discussion about each of the 14 regions at the expense of individual residue-specific changes. The B-factors associated with each amino acid were extracted from the PDB file

1DWR and then averaged for the regions listed above to allow comparisons of B-factors to the fragmentation yields. GetArea was also used to obtain backbone and side-chain solvent accessible surface areas.[94]

2.3 Results:

The dominant species observed in the native ESI mass spectrum of both holo- and apo-myoglobin have low charge states (8+ to 10+), unlike the much higher charge states favored by myoglobin under standard denaturing conditions (see **Figure 2.1**). The ions in low charge states are characteristic of native-like structures, as documented in previous mass spectrometric studies.[61–64,91,95] For both holo and apo forms the most abundant charge state is 9+. For the holo (heme-containing) protein there is little to no apo form observed in the native ESI mass spectrum. However, for the apo form low abundance ions in higher charge states are observed (**Figure 2.1b**) which represent unfolded species. Additionally, some heme-containing species of moderate abundance are observed in the 8+ and 9+ charge states due to incomplete removal of the heme group during preparation of apo-myoglobin. Because the holo and apo forms of myoglobin differ in mass by more than 600 Da, these species are easily resolved and well isolated from each other during subsequent MS/MS analysis. The 8+ and 9+ charge states were determined to have more native-like cross-sections when sprayed between pH 7.0 and 4.0 using native ESI conditions in a recent study by Vahidi *et al*.[91] Each of the low "native" charge states of holo- and apo-myoglobin (i.e. 8+, 9+, 10+) was isolated and activated separately using UVPD. Comparison of the UVPD mass spectra of holo and apo-myoglobin (9+) is shown in **Figure 2.2**. Expansion of one higher m/z region shows the mass shift associated with retention of the heme group for some of the *a* and *x* ions produced by holo-myoglobin, demonstrating the ability to track these non-covalent interactions. Such heme-containing ions were not found upon HCD.

The resulting sequence coverages obtained by UVPD for the 8+, 9+, and 10+ charge states were 53%, 87%, and 88% for holo-myoglobin and 64%, 86% and 84% for apo-myoglobin, respectively. The retention of the heme did not notably influence the total sequence coverage obtained. For both holo- and apo-myoglobin there was a shift (an increase) in total sequence coverage from 8+ to 9+, which may be the result of a conformational difference between the charge states in the gas phase or related to the availability of one more proton for charge-mediated fragmentation pathways. The sequence coverage obtained by UVPD of the 20+ species of the denatured protein (heme free) was 77%. Interestingly, the corresponding ETD sequence coverages were low for myoglobin in the native-like charge states (e.g. a sequence coverage of only 18% was attained for 10+). The fragmentation yields were more substantial upon ETD of the denatured 20+ state, giving a sequence coverage of 70% which reflects the well-known charge state dependence of electron activation methods. It has been reported that many salt bridges are not disrupted upon electron activation of native protein structures in low charge states, thus depressing the total sequence coverage of ETD.[71] The low overall sequence coverage in the present study confounded deeper correlation of ETD fragmentation with specific structural elements of native states of myoglobin, and thus the ETD results are not discussed further. Each MS/MS spectrum was processed to calculate the normalized fragmentation yield per amino acid (i.e. based on backbone cleavages between every pair of residues) as described in the experimental section. These values were plotted with their associated standard deviations in **Figure 2.3** which displays the normalized fragment yield for each native charge state (8+, 9+, 10+) for both forms of myoglobin. The results for the denatured charge state (20+) are included in the apo-myoglobin panel.

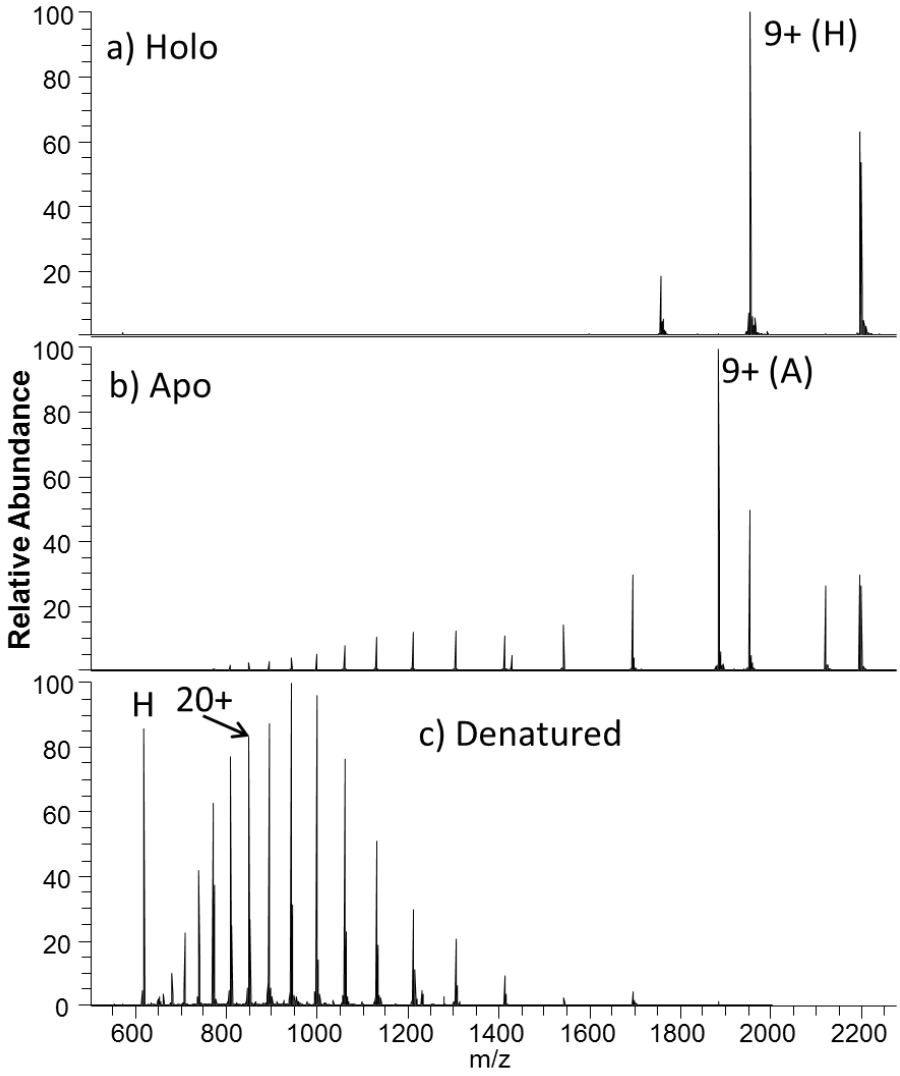


Figure 2.1 a) Holo-myoglobin and b) apo-myoglobin sprayed under native conditions (5mM ammonium acetate at pH 5.5) at 1.5-1.7 kV and a capillary temperature of 150°C. c) denatured myoglobin in 49.9% water, 50% acetonitrile and 0.1% formic acid sprayed at 4kV with a 300°C capillary temperature. (H) stands for HOLO, (A) stands for APO and H denotes the free heme group.

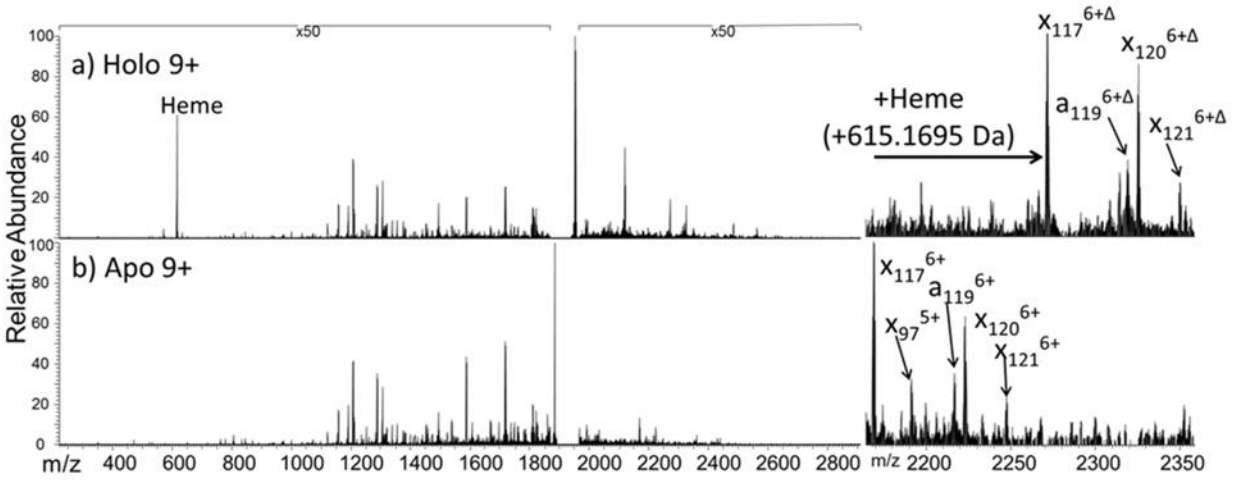


Figure 2.2 193 nm UVPD mass spectra of the 9+ charge state of the a) holo and b) apo form of myoglobin. Panels (c,d) to the right are expanded regions (m/z 2175 to 2350) with several ions labelled to illustrate the mass shift (+615.1695 Da) affiliated with retention of the heme group. Δ indicates the inclusion of heme in the fragment ions.

The general shapes of the fragment distributions in **Figure 2.3** are similar for the holo- and apo-forms, with clusters of backbone sites showing elevated fragmentation yields relative to others that exhibit little or no yields. The graphs in **Figure 2.3** also include demarcated regions which are defined based on fourteen known structural regions of myoglobin. If the protein cleaved uniformly at each backbone position, the expected fragmentation yield would be 0.66% (e.g., 100% fragment ion current divided by 152 backbone cleavage sites). Backbone cleavages between residues that were detected to give greater than 0.66% fragmentation yield are considered to be residues exhibiting preferential (enhanced) cleavage. These specific amino acids are highlighted in the primary sequence map (bottom of **Figure 2.4**) and are also superimposed on the myoglobin crystal structure (PDB code 1DWR), as illustrated for each native holo-myoglobin charge state in **Figure 2.4**. In an effort to streamline the data interpretation in a more structurally meaningful manner, the backbone site-specific fragment yields were grouped and

averaged according to the fourteen known structural regions of myoglobin and are displayed in histogram format in **Figure 2.5** for holo-myoglobin (with structural regions labelled in **Figure 2.6)** for both HCD and UVPD. The analogous results for apo-myoglobin and denatured myoglobin are shown in **Figure 2.7.** Essentially this treatment provides fragment yield densities which are a metric of fragmentation localized to specific structural regions of the protein, normalized for the size of the region. This provides a way to visualize the change in fragmentation yield for each activation method and as a function of charge state. To facilitate comparisons, each structural region is assigned an average B-factor value (as described in the Experimental Section), which conveys the degree of flexibility in the crystal structure. The B-factors for each defined region of holo-myoglobin (1DWR) are shown in **Figure 2.5**.

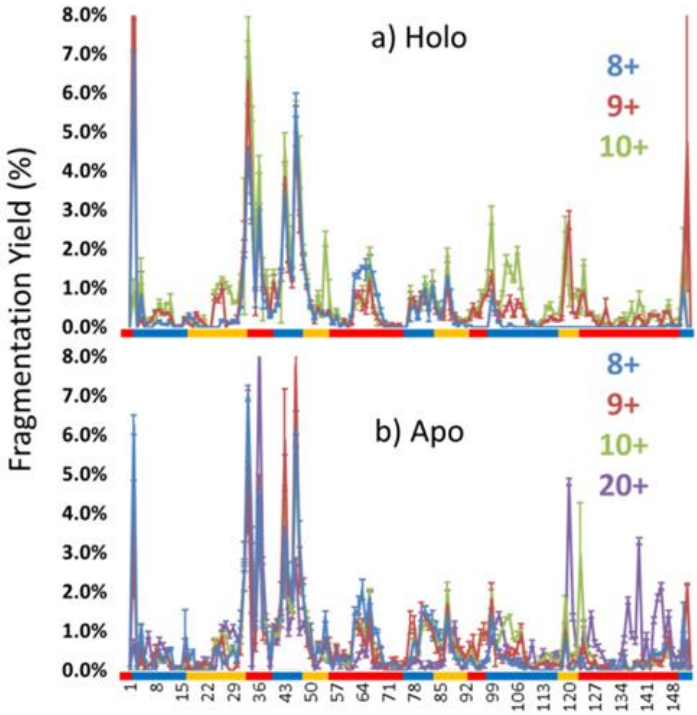


Figure 2.3 Normalized fragmention yields upon UVPD of the 8+ to 10+ charge states for both a) holo and b) apo forms based on averaging 3-4 replicates. The fragmentation yields for the 20+ denatured state of apo-myoglobin is shown in b as well. Structural regions are demarcated below the graphs with colored bars showing the loops and helices.

The sequence or order of unfolding of myoglobin based upon molecular dynamics simulations, CD experiments, NMR, and HDX mass spectrometric data provides another perspective on the structural stability of the protein.[41,42,44,89,96] As briefly summarized here based on published findings,[41,42,44,89,96] the F-helix is the most unstable structural element, thus the first to unfold. This is followed by unfolding of the C,D, and E helices, prior to the unfolding of the B-helix. The last helices to unfold are the A,G, and H helices or the AGH core. Not surprisingly, this AGH core is found to be consistently the most stable and least flexible, with the exception of the N-terminus of the G-helix, which has been shown by computational modeling to fray.[89] B-factors from the 1DWR crystal structure were also analyzed and used to correlate the fragmentation yields in the present study to a measured and quantified flexibility value (B-factor). The averaged B-factors of the previously defined demarcated regions are summarized in **Figure 2.8a**. Those regions with high B-factor values are flexible regions of the protein, while those with low B-factor values are more rigid and more stable. The charge state-based fragmentation yield data obtained from UVPD of both the holo and apo forms is shown in **Figure 2.8b** and **2.8c,** respectively. The UVPD results obtained for the apo-form should not be compared directly to the B-factor values because the latter includes the impact of the heme group. However, comparison of the B-factor values with the UVPD data for holo-myoglobin and comparison of the UVPD data for the holo and apo proteins provides interesting insight, as described in this study. A more detailed comparison of the UVPD results for the apo and holo forms as a function of charge state is shown in **Figure 2.11**, also discussed in detail later.

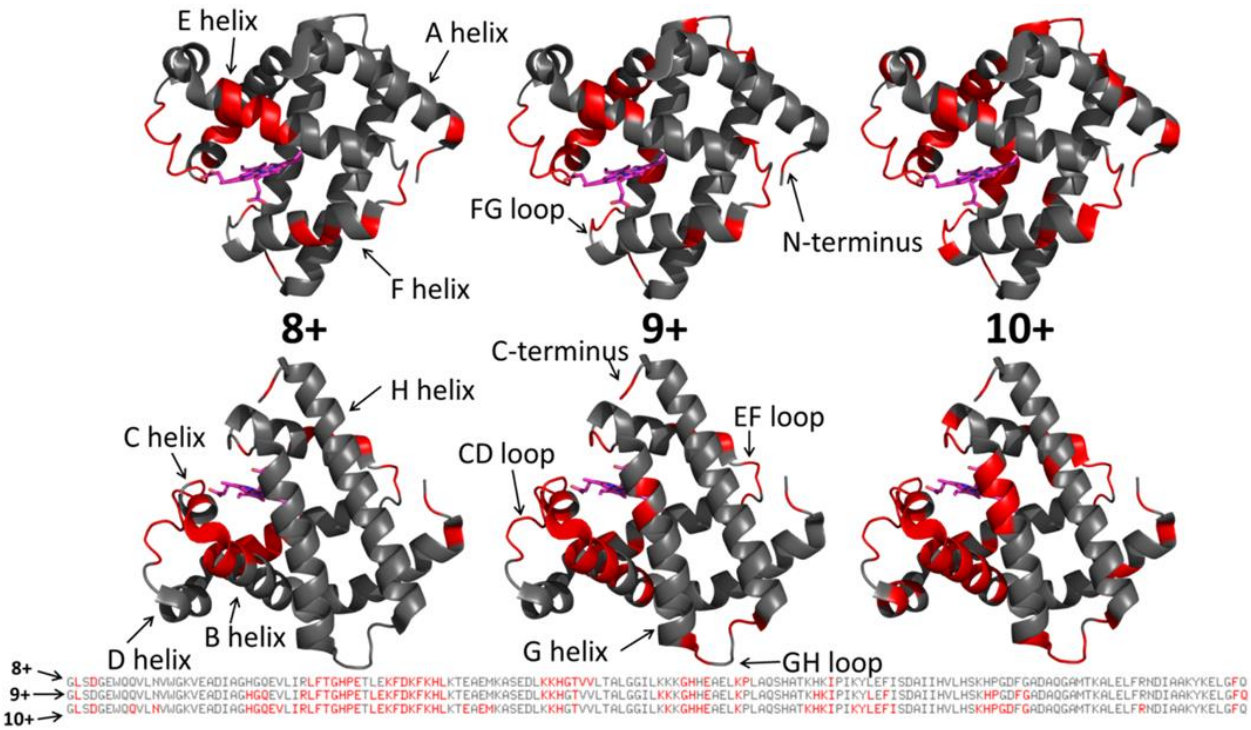


Figure 2.4 Those residues that were found to have enhanced backbone cleavage frequencies upon UVPD activation are highlighted in red on the holo-myoglobin crystal structure 1DWR, as also indicated in the primary sequence below for the native charge states (8+, 9+, 10+) evaluated. The heme group is shown as a hot pink color. The 1DWR crystal structure has been rotated 180° about the x-axis between the top and bottom structures.

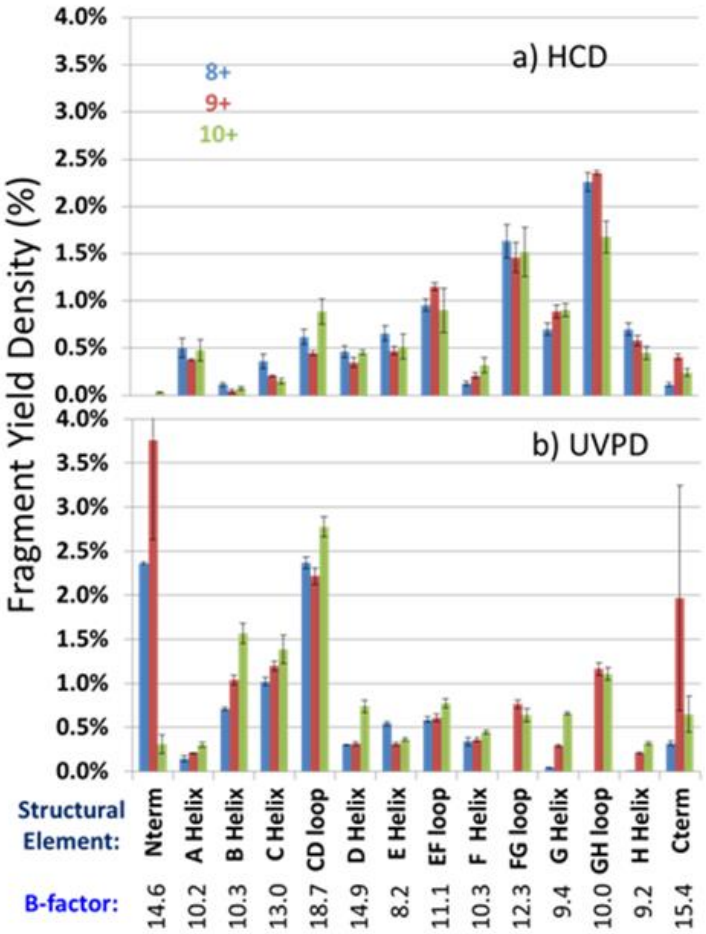


Figure 2.5 Normalized fragmentation yields averaged over the total number of amino acids in each structural element for charge states 8+ (blue bars), 9+ (red bars), and 10+ (green bars) of holo-myoglobin with (a) HCD and (b) UVPD. Numbers under the structural regions show the calculated average B-factors for holo-myoglobin.

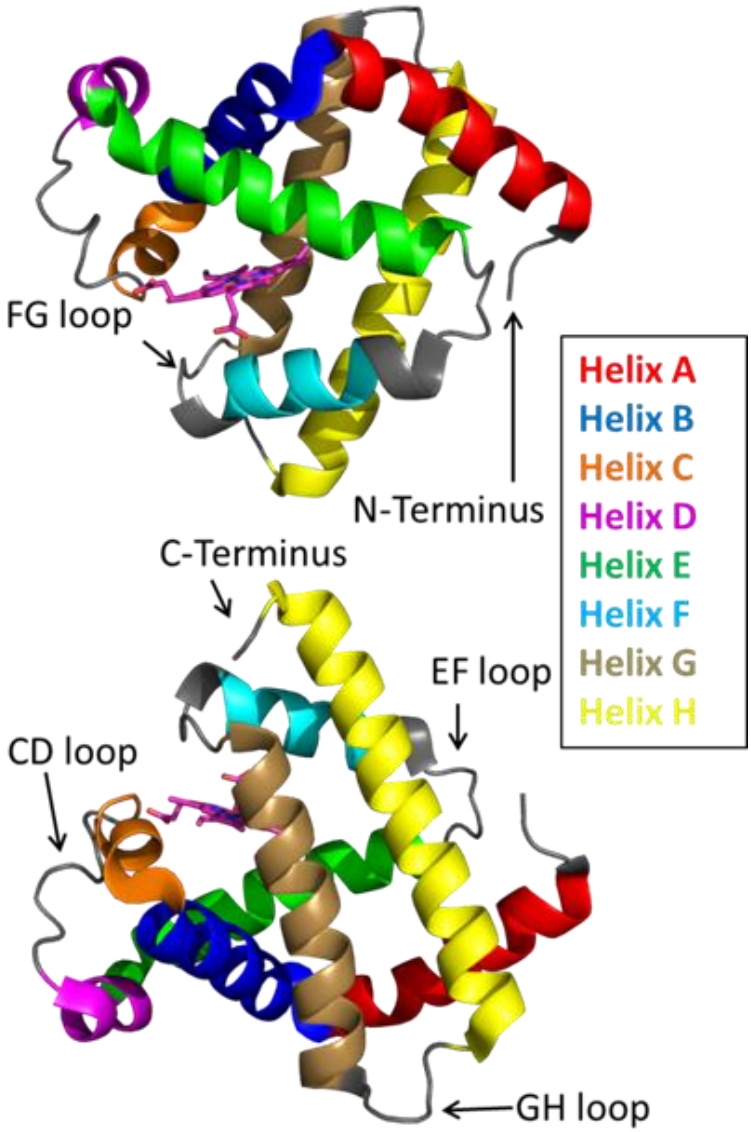


Figure 2.6 A schematic representation of holo-myoglobin with color coded helices, loops and termini, and the heme is shown in hot pink. The 1DWR crystal structure has been rotated 180° about the x-axis between the top and bottom structures.

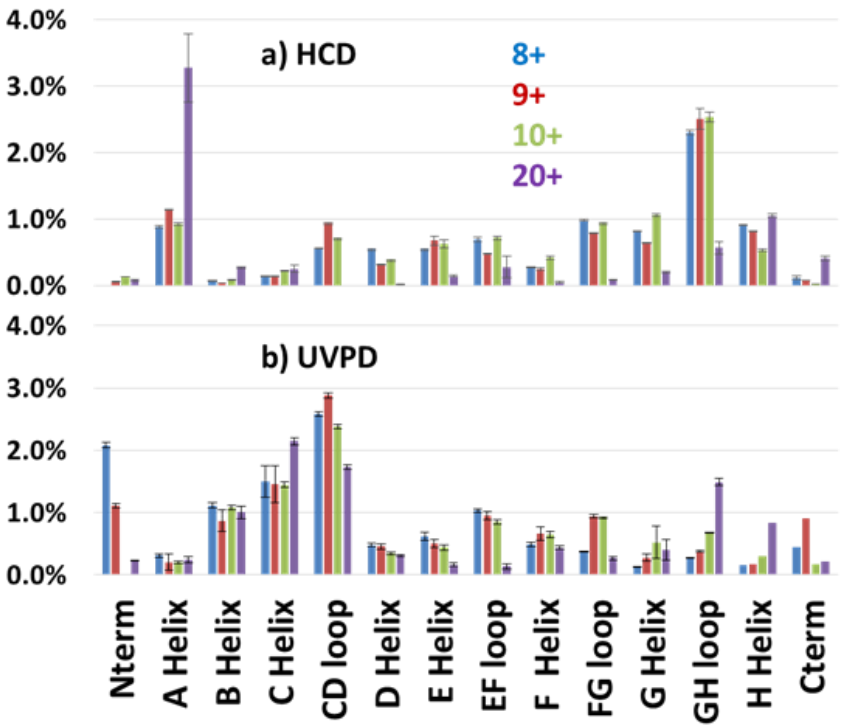


Figure 2.7 Normalized fragmentation yields averaged over the total number of amino acids in each structural element for charge states 8+ (blue bars), 9+ (red bars), 10+ (green bars) and 20+ (purple bars) of apo-myoglobin with HCD (a) and UVPD (b) fragmentation.

For the present study, our goal was to assess whether the UVPD fragmentation yields (**Figure 2.3** and **2.5**) reflected the local stabilities of the tertiary and secondary structural features of myoglobin (i.e. structures in **Figure 2.4**) as suggested previously in a number of publications utilizing electron activation and 266 nm UVPD for other proteins.[22,79,81,97] In essence, these previous studies indicated that those regions of a protein that displayed high fragmentation frequencies and thus resulted in high fragment yields were presumed to be conformationally flexible and/or have few retained intramolecular interactions. Those which had low fragmentation yields were conformationally stable elements, such as those with strong intramolecular interactions in the gas phase. A number of substantial variations in relative fragmentation yields are observed in **Figure 2.8** as a function of charge state and apo/holo state of myoglobin (**Figure 2.9**).

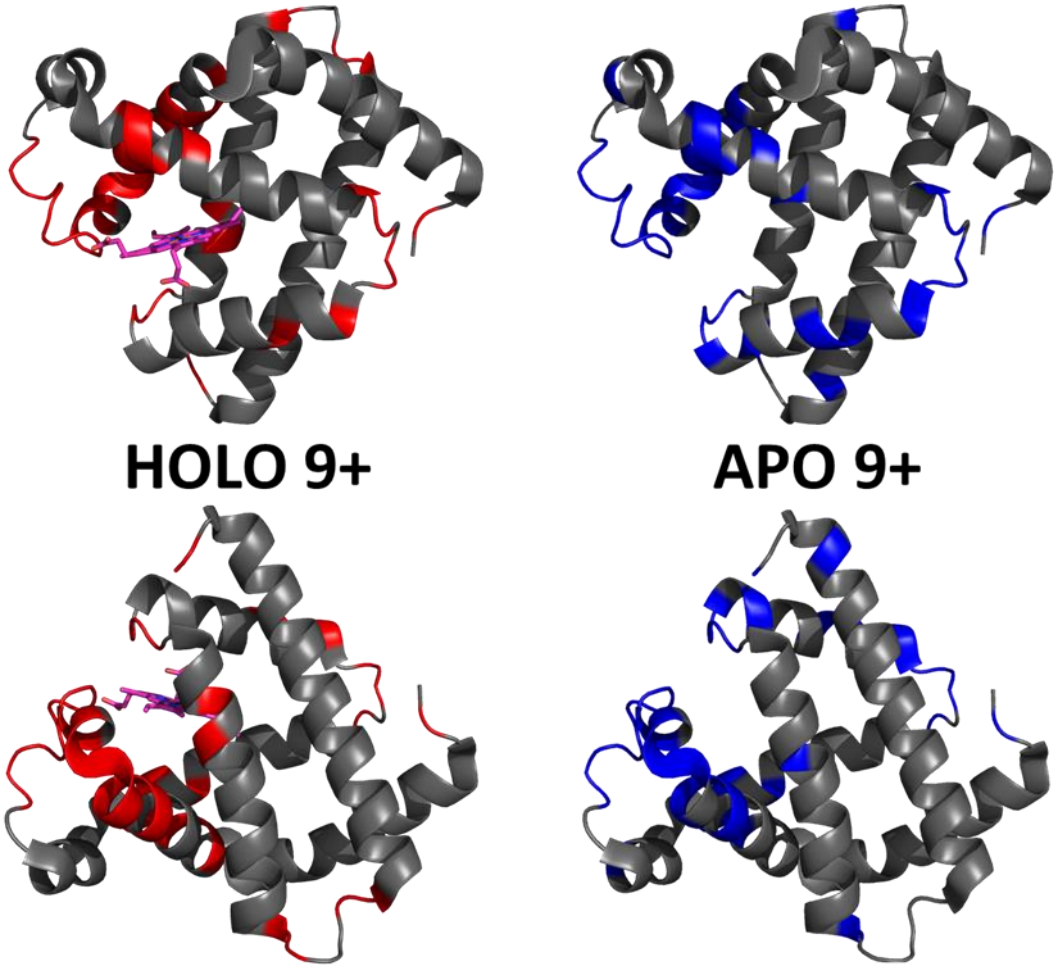


Figure 2.9 Those residues that were found to have enhanced backbone cleavage frequencies upon UVPD activation are highlighted in red on the holo-myoglobin crystal structure 1DWR and in blue on the apo-myoglobin structure. The heme group is shown as a hot pink color for holo-myoglobin. The 1DWR crystal structure has been rotated 180° about the x-axis between the top and bottom structures. The holo-myoglobin crystal structure is used for the apo-protein to allow easy visualization of the regions that show enhanced UVPD fragmentation.

2.4 Discussion

2.4.1 UVPD of Holo-myoglobin

Most of the preferential fragmentation upon UVPD of the three native charge states of both holo- and apo-myoglobin occurred in the loop regions (**Figure 2.5b** and **Figure 2.7b**), which typically have larger B-factors compared to helices. There appears to be good agreement between both the B-factors and the sequence of unfolding in solution with the 193 nm UVPD results, particularly with the lowest charge state (8+) interrogated (**Figure 2.8**). For the 8+ charge

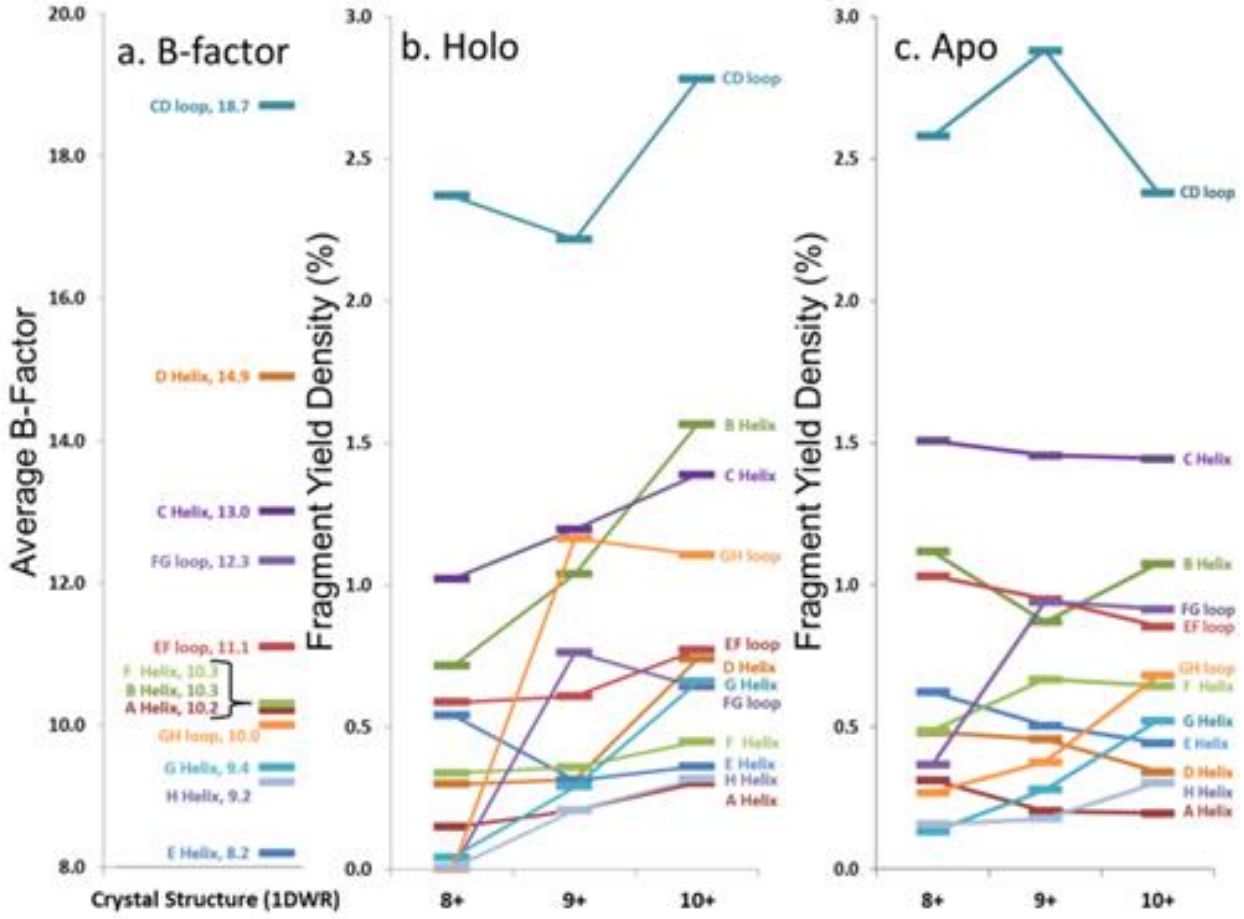


Figure 2.8 Comparison of average (a) B-factor value s and fragment yield densities of respective labelled regions from the (b) holo and (c) apo forms of myoglobin for the three native charge states. Regions with greater B-factors or fragmentation yield densities indicate less stable regions whereas those with lower values indicate more stable regions. The C- and N-termini have been removed from the figure. Standard deviations for fragment yield densities are found in Figure 4 and Figure S2 for holo and apo forms, respectively.

state of holo-myoglobin, the AGH core helices exhibited the lowest fragmentation yields and in fact mirrored the order of stability predicted from the B-factors. The GH loop also has one of the lowest UVPD fragmentation yields, again consistent with the low B factor value. At the other extreme, the CD loop and C helix had the highest UVPD fragmentation yields, outcomes also reflected in the B factors which are two of the highest among all the structural elements. There are other structural elements for which the UVPD fragmentation yields and B values did not agree so well. These include the FG loop and D helix which had low UVPD fragmentation yields but relatively high B values, and the E helix exhibited a moderate fragmentation yield but had the lowest B value of all structural elements.

When comparing the UVPD fragment yields obtained for the 8+ charge state to the 9+ charge state, fragmentation of the FG and GH loop regions was greatly enhanced, along with significant increases in the fragmentation of the connecting helices G and H (in all cases going from virtually no detectable fragmentation for the 8+ charge state to moderate/significant fragmentation for the 9+ charge state) (**Figures 2.5b** and **2.8b**). The fragmentation yields of the B and C neighboring helices also increased with charge state. The CD loop and the E helix are two regions that showed moderate decreases in fragmentation going from the 8+ to 9+ charge state, and several other structural regions, including the A,D and F helices and the EF loop, exhibited no significant change as a function of charge state. Collective grouping of these regions (those showing decreasing fragmentation or no change with charge state) reveals that all are important for heme binding.[98] It has also been noted that absorption cross-sections of proteins change with charge density and conformation.[99] This prior finding might explain in part some of the changes observed in the UVPD fragmentation yields as a function of charge state in the present study, namely a possible charge-density induced change in absorption cross-section in

addition to a degree of protein unfolding as the charge state increases. Upon transitioning to the 10+ charge state, the most stable regions predicted by UVPD (e.g. lowest fragmentation yields) are the A, H, E and F helices, all of which increased slightly from the fragmentation yields determined for the 9+ charge state. The G helix displays a more significant increase in fragmentation yield for the 10+ charge state, an outcome attributed to N-terminal fraying as a consequence of unfolding of the FG loop that was also noted from the 8+ to 9+ charge states. This type of fraying has been noted in prior solution studies.[89,96] Several of the structural regions show an increase in fragmentation for the 10+ charge state, suggestive of unfolding related to the D, G, C, and B helices along with EF and CD loops. This nearly global increase in fragmentation with charge state tracks well with the concept of electrostatic-induced elongation. Even with this alpha helical fraying and expansion in the regions from B to the D helix, the maintenance of the E and F helices as well as the AGH core of the protein allow retention of interactions with the heme group.

Solvent accessibility is another physical parameter that has shown positive correlation with B-factors of proteins[100,101] as well as with ETD fragmentation trends for native proteins,[70] and thus was also evaluated in the context of the UVPD data in the present study. A plot of the solvent accessible surface area (SASA) for both the backbone and side chains generated from the 1DWR crystal structure is shown in **Figure 2.10** (along with the B-factor values and UVPD fragmentation yields for ready comparison). There is some general correlation backbone SASA and the UVPD fragmentation yield for those regions having SASA above 50%. Those regions that have both low SASA values for the side-chain and backbone regions also have the lowest UVPD fragmentation yields, particularly for the 8+ charge state. The A-helix region shows the worst correlation between SASA and UVPD fragmentation yields, with rather prominent SASA

values for the side-chains of the A-helix region of the protein but yet very low UVPD yields. However, the SASA values of the backbone in this A-helix region are very low.

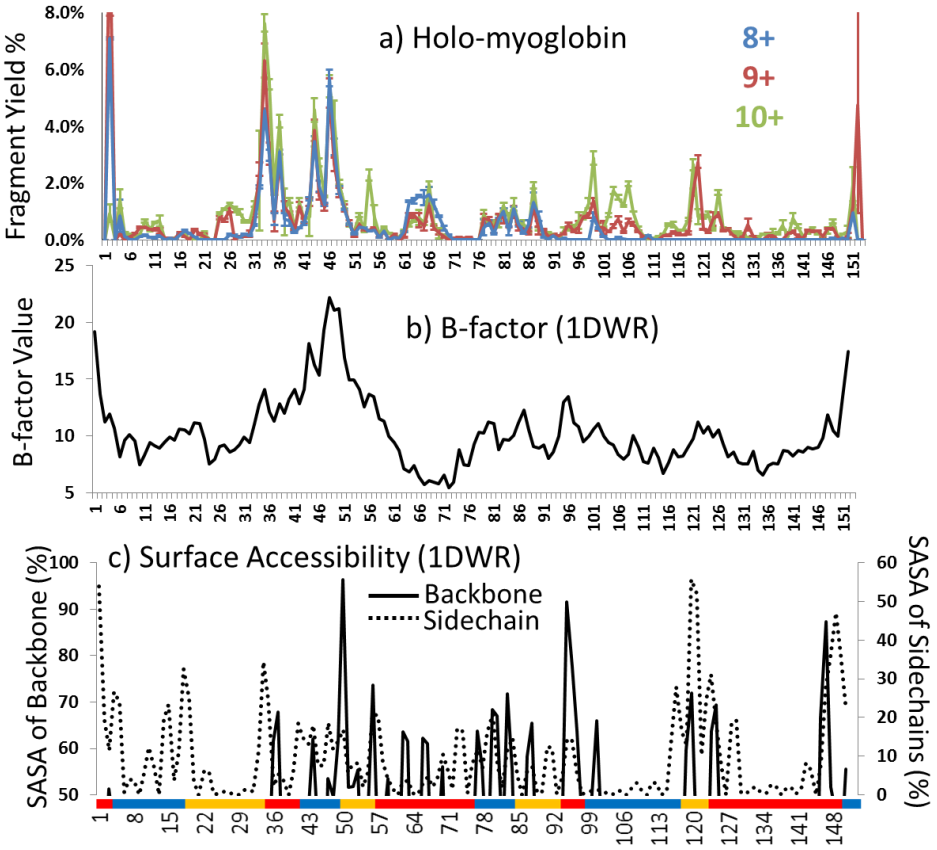


Figure 2.10 a) Holo-myoglobin fragmentation yields as reported in Figure 2 are presented for comparison with (b) the B-factors from holo-myoglobin crystal structure 1DWR along with (c) surface accessibility data of both the backbone and side-chains calculated with GetArea. Regions are demarcated below (c). Surface accessibility data was smoothed by box car averaging of two points.

The observation that the loop regions tend to fragment more readily than the helices by UVPD may also be related in part to the larger molar absorptivities of loops compared to helices at 193 nm. Tsai *et al.* reported that the λ_{max} of a typical alpha-helix occurred between 189-204 nm in solution with a molar absorptivity of 4.1 cm^2*mol x 10^{-3}, whereas the molar absorptivities of coiled or loop regions were around 6.9 cm^2*mol x 10^{-3} with a similar λ_{max} (187-190 nm).[102] Interestingly, beta-sheets were reported to have an even greater molar absorptivity;[102] however,

myoglobin contains no beta-sheets.[102] Although the exact mechanism of the UV activation and dissociation process for large molecules like proteins remains unknown in the gas phase, it is conceivable that local absorptivities of certain structural elements of proteins that are retained in the gas phase could have an impact on energy deposition and re-distribution prior to fragmentation. More general trends about fragmentation efficiencies of helical and non-helical regions of native proteins may emerge as additional proteins are analyzed by UVPD.

2.4.2 UVPD fragmentation behavior of Holo- versus Apo-myoglobin

Apo-myoglobin adopts a structure similar to holo-myoglobin except with greater disorder and partial unfolding of the F helix, adjacent EF and FG loops, N-terminus of the G helix and H helix.[103,104] UVPD fragmentation trends were constructed for apo-myoglobin (produced via chemically-induced removal of heme in solution, followed by native ESI of the heme-free protein). The normalized fragmentation yields for the backbone-specific cleavages are shown in **Figure 2.3b**, and the averaged fragmentation yields grouped by the 14 structural elements are illustrated in **Figure 2.7b** and tracked versus charge state in **Figure 2.8c**. The trends for the holo- versus apo- forms of myoglobin are summarized in **Figure 2.11** for the three native charge states, thus allowing a more detailed assessment of the impact of the heme ligand. Previous circular dichroism (CD) analysis of holo and apo myoglobin has shown that there is a reduction in alpha helical content for the apo form relative to the holo form,[105–107] so this feature could account for some variations in the observed UVPD fragmentation yields. For example, virtually all regions of the heme-free protein have greater fragmentation yields than the holo-form for the 8+ charge state, an outcome that is consistent both with the CD and spectroscopic solution data and the differences in molar absorptivity attributed to the increasing loop and decreasing alpha helical content. Relative to holo-myoglobin, the most dramatic increases in fragmentation

density for apo-myoglobin occur for the B and C helices as well as the EF loop regions, followed by less significant increases in fragmentation yields in the CD, FG and GH loops. The larger changes in fragmentation density in these regions may be indicative of unfolding or partial unfolding events in regions which were stabilized by the heme group binding for the holo form of the protein. Aside from these increases in fragmentation density or unfolding for the holo form, the AGH core and FH/GH loops still exhibit the lowest fragmentation yields for apo-myoglobin.

One might anticipate a general across-the-board increase in UVPD fragmentation yields for apo-myoglobin relative to holo-myoglobin because the collisional cross-section of myoglobin increases, albeit slightly, upon transition from the holo to myo form.[14] Although there are modest increases in UVPD yields for many of the structural elements (AGH helical core, D,E,F helices) going from the holo to myo form, a few regions (EF loop, B and C helices, CD, EF and FG loops) show more significant increases. The increases in the C, D, E and F helix fragmentation yields for the apo form may reflect the lack of stabilizing non-covalent interactions with the heme group that also allows the intervening loop regions to release and helices to unfold. For the 9+ and 10+ charge states, two regions showed a decrease in UVPD fragmentation for the apo form: the B helix and GH loop, as well as the CD loop and D helix for the 10+ charge state. The substantial amount of change in fragmentation yields for the 10+ charge state of holo- versus apo-myoglobin may reflect significant differences in secondary and possibly tertiary structures. The agreement in the UVPD trends between the holo and apo forms of myoglobin with a particular strong similarity in fragmentation behavior of the AGH core helices suggesting that this helical bundle is stabilized in both of these forms.

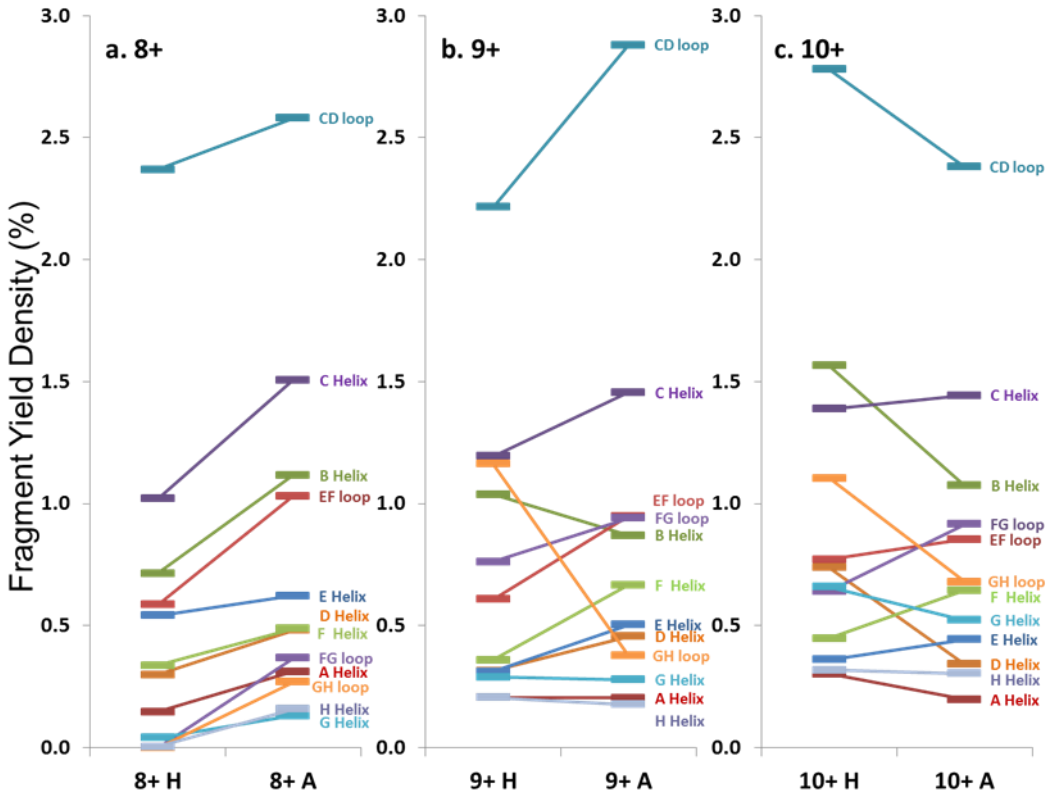


Figure 2.11. Direct comparison of fragmentation yield densities for holo and apo myoglobin : (a) 8+, (b) 9+, (c) 10+. The key structural elements are labelled. The N- and C-terminii have been removed from the figure.s Standard deviations for fragment yield densities are shown in Figure 4 and Figure S2 for holo and apo forms, respectively.

2.4.3 HCD of Holo and Apo-myoglobin

A final comparison was undertaken to evaluate the fragmentation trends of myoglobin obtained by HCD, another activation method that has shown promise for characterization of intact proteins in top-down studies. To date, HCD has not been reported to cause fragmentation that is specific for native protein structures, but in general it has not yet been extensively explored.[62,82] The HCD fragmentation results are shown for both holo and apo forms in **Figure 2.5a** and **Figure 2.7a,** respectively, and exhibit similar trends. Each displays modest variation in the fragmentation yields among the three lowest charge states and a far more drastic change for the elongated protein in the 20+ charge state. For the latter 20+ charge state, fragmentation was

mainly localized near the termini of the protein (A- and H-helices, GH loop), an outcome that has been noted previously for HCD of a variety of proteins in high charge states (i.e. the type of denatured proteins commonly analyzed in top-down proteomics studies). HCD of the native-like charge states showed more extensive fragmentation of the EF and FG loop regions in holo-myoglobin compared to UVPD. In many cases, the regions that give the lowest fragmentation yields by HCD were not those with the lowest B factors (most stable) nor did the trend mirror the established sequence of protein unfolding. For example the AGH helices, which are known to be the most stable core elements of myoglobin, exhibited among the highest fragmentation yields upon HCD. The GH loop is considered the most stable loop element in light of its low B-factor value (10.0), yet its HCD fragmentation yield was very high. The B, C, D and F helices displayed the lowest fragmentation yields upon HCD which opposes the trend in the order of unfolding. Interestingly, several of the regions that exhibited the most extensive fragmentation upon UVPD (B,C helices and CD loop) had among the lowest HCD yields, and similarly the AGH helical core and GH loop were found to exhibit substantial fragmentation upon HCD but not upon UVPD. The lack of consistent correlation between the HCD fragmentation yields and the order of solution unfolding and B-factors, suggests that the impact of the mobile protons which facilitate collisional activated backbone cleavages and residue-specific preferential cleavages (i.e. those affiliated with the presence of Asp, Glu, Pro, Gly, Ile, Leu and Val residues) substantially modulates HCD fragmentation in the gas phase, even for those proteins assumed to be in native-like conformations (low charge states).

2.5 Conclusions

For the low charge states of myoglobin (presumed native-like structures), the trend in the UVPD fragmentation yields (from largest yield to lowest yield) showed good agreement with the

average B-factor trend (from highest B-factor to lowest), especially for the 8+ and 9+ states of holo-myoglobin. The UVPD fragmentation trends appear to reflect little dependence on side-chain interactions and rather on secondary and tertiary interactions with amide hydrogens. For example, in a folded helix the backbone amide-hydrogens are involved in strong hydrogen bonding interactions, whereas in a loop region these amide-hydrogens are not heavily involved in stabilizing interactions. The corresponding UVPD fragment yields suggested that cleavages were favored at backbone positions for which the amides were not involved in hydrogen bonding interactions. As the charge state increased, the UVPD fragmentation yields increased for the N-terminus region of the G-helix, the F-helix and the FG loop. Additionally, UVPD led to preferential backbone cleavages in regions of the protein with higher relative B-factors, such as loop regions and frayed helical ends. The increase in fragment yields for specific structural regions upon UVPD of the 8+ to 10+ states of myoglobin are consistent with unfolding of the G and F helices and FG and GH loops. Differences in fragmentation may also arise in part from the difference in molar absorptivities between the various secondary structures in which loops have greater absorptivities than helices. Comparison of UVPD fragmentation yields for the holo and apo forms revealed similarities in the AGH helical core of the protein, whereas the fragmentation yields were suppressed in holo-myoglobin for those regions known to participate in heme binding. HCD displayed enhanced backbone fragmentation in regions that were more likely reflective of the known preferential charge-directed residue-specific backbone cleavages affiliated with collisional activation rather than correlation with the least stable regions based on B-factors.

Bibliography

(1) Yu, H. Extending the Size Limit of Protein Nuclear Magnetic Resonance. *Proc. Natl. Acad. Sci.* **1999**, *96*, 332–334.

(2) Acharya, K. R.; Lloyd, M. D. The Advantages and Limitations of Protein Crystal Structures. *Trends Pharmacol. Sci.* **2005**, *26*, 10–14.

(3) Konermann, L.; Tong, X.; Pan, Y. Protein Structure and Dynamics Studied by Mass Spectrometry: H/D Exchange, Hydroxyl Radical Labeling, and Related Approaches. *J. Mass Spectrom.* **2008**, *43*, 1021–1036.

(4) Fitzgerald, M. C.; West, G. M. Painting Proteins with Covalent Labels: What's In the Picture? *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 1193–1206.

(5) Mendoza, V. L.; Vachet, R. W. Probing Protein Structure by Amino Acid-Specific Covalent Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* **2009**, *28*, 785–815.

(6) Xu, Y.; Schmitt, S.; Tang, L.; Jakob, U.; Fitzgerald, M. C. Thermodynamic Analysis of a Molecular Chaperone Binding to Unfolded Protein Substrates. *Biochemistry* **2010**, *49*, 1346–1353.

(7) Xu, Y.; Falk, I. N.; Hallen, M. A.; Fitzgerald, M. C. Mass Spectrometry- and Lysine Amidination-Based Protocol for Thermodynamic Analysis of Protein Folding and Ligand Binding Interactions. *Anal. Chem.* **2011**, *83*, 3555–3562.

(8) Tong, X.; Wren, J. C.; Konermann, L. Γ-Ray-Mediated Oxidative Labeling for Detecting Protein Conformational Changes by Electrospray Mass Spectrometry. *Anal. Chem.* **2008**, *80*, 2222–2231.

(9) Zhang, H.; Gau, B. C.; Jones, L. M.; Vidavsky, I.; Gross, M. L. Fast Photochemical Oxidation of Proteins (FPOP) for Comparing Structures of Protein/Ligand Complexes: The Calmodulin-Peptide Model System. *Anal. Chem.* **2011**, *83*, 311–318.

(10) Gau, B. C.; Sharp, J. S.; Rempel, D. L.; Gross, M. L. Fast Photochemical Oxidation of Protein Footprints Faster than Protein Unfolding. *Anal. Chem.* **2009**, *81*, 6563–6571.

(11) Stocks, B. B.; Sarkar, A.; Wintrode, P. L.; Konermann, L. Early Hydrophobic Collapse of α1-Antitrypsin Facilitates Formation of a Metastable State: Insights from Oxidative Labeling and Mass Spectrometry. *J. Mol. Biol.* **2012**, *423*, 789–799.

(12) Hambly, D.; Gross, M. Laser Flash Photochemical Oxidation to Locate Heme Binding and Conformational Changes in Myoglobin. *Int. J. Mass Spectrom.* **2007**, *259*, 124–129.

(13) Zhang, H.; Shen, W.; Rempel, D.; Monsey, J.; Vidavsky, I.; Gross, M. L.; Bose, R. Carboxyl-Group Footprinting Maps the Dimerization Interface and Phosphorylation-Induced Conformational Changes of a Membrane-Associated Tyrosine Kinase. *Mol. Cell. Proteomics* **2011**, *10*.

(14) Zappacosta, F.; Ingallinella, P.; Scaloni, A.; Pessi, A.; Bianchi, E.; Sollazzo, M.; Tramontano, A.; Marino, G.; Pucci, P. Surface Topology of Minibody by Selective Chemical Modifications and Mass Spectrometry. *Protein Sci. Publ. Protein Soc.* **1997**, *6*, 1901–1909.

(15) Suckau, D.; Mak, M.; Przybylski, M. Protein Surface Topology-Probing by Selective Chemical Modification and Mass Spectrometric Peptide Mapping. *Proc. Natl. Acad. Sci.* **1992**, *89*, 5630–5634.

(16) Glocker, M. O.; Borchers, C.; Fiedler, W.; Suckau, D.; Przybylski, M. Molecular Characterization of Surface Topology in Protein Tertiary Structures by Amino-Acylation and Mass Spectrometric Peptide Mapping. *Bioconjug. Chem.* **1994**, *5*, 583–590.

(17) Izumi, S.; Kaneko, H.; Yamazaki, T.; Hirata, T.; Kominami, S. Membrane Topology of Guinea Pig Cytochrome P450 17α Revealed by a Combination of Chemical Modifications and Mass Spectrometry. *Biochemistry* **2003**, *42*, 14663–14669.

(18) Turner, B. T.; Sabo, T. M.; Wilding, D.; Maurer, M. C. Mapping of Factor XIII Solvent Accessibility as a Function of Activation State Using Chemical Modification Methods†. *Biochemistry* **2004**, *43*, 9755–9765.

(19) O'Brien, J. P.; Pruet, J. M.; Brodbelt, J. S. Chromogenic Chemical Probe for Protein Structural Characterization via Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 7391–7397.
(20) Makoff, A. J.; Malcolm, A. D. Properties of Methyl Acetimidate and Its Use as a Protein-Modifying Reagent. *Biochem. J.* **1981**, *193*, 245–249.
(21) Janecki, D. J.; Beardsley, R. L.; Reilly, J. P. Probing Protein Tertiary Structure with Amidination. *Anal. Chem.* **2005**, *77*, 7274–7281.
(22) Zhou, Y.; Vachet, R. W. Increased Protein Structural Resolution from Diethylpyrocarbonate-Based Covalent Labeling and Mass Spectrometric Detection. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 708–717.
(23) Jumper, C. C.; Bomgarden, R.; Rogers, J.; Etienne, C.; Schriemer, D. C. High-Resolution Mapping of Carbene-Based Protein Footprints. *Anal. Chem.* **2012**, *84*, 4411–4418.
(24) Jumper, C. C.; Schriemer, D. C. Mass Spectrometry of Laser-Initiated Carbene Reactions for Protein Topographic Analysis. *Anal. Chem.* **2011**, *83*, 2913–2920.
(25) Liu, X.; Broshears, W. C.; Reilly, J. P. Probing the Structure and Activity of Trypsin with Amidination. *Anal. Biochem.* **2007**, *367*, 13–19.
(26) Hunter, M. J.; Ludwig, M. L. The Reaction of Imidoesters with Proteins and Related Small Molecules. *J. Am. Chem. Soc.* **1962**, *84*, 3491–3504.
(27) Running, W. E.; Reilly, J. P. Variation of the Chemical Reactivity of Thermus Thermophilus HB8 Ribosomal Proteins as a Function of pH. *PROTEOMICS* **2010**, *10*, 3669–3687.
(28) Running, W. E.; Reilly, J. P. Ribosomal Proteins of Deinococcus Radiodurans: Their Solvent Accessibility and Reactivity. *J. Proteome Res.* **2009**, *8*, 1228–1246.
(29) Beardsley, R. L.; Running, W. E.; Reilly, J. P. Probing the Structure of the Caulobacter Crescentus Ribosome with Chemical Labeling and Mass Spectrometry. *J. Proteome Res.* **2006**, *5*, 2935–2946.
(30) Liu, X.; Reilly, J. P. Correlating the Chemical Modification of Escherichia Coli Ribosomal Proteins with Crystal Structure Data. *J. Proteome Res.* **2009**, *8*, 4466–4478.
(31) Running, W. E.; Ni, P.; Kao, C. C.; Reilly, J. P. Chemical Reactivity of Brome Mosaic Virus Capsid Protein. *J. Mol. Biol.* **2012**, *423*, 79–95.
(32) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 806–812.
(33) Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. Hydrogen/Deuterium Exchange Mass Spectrometry with Top-Down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 kDa Protein. *J. Am. Chem. Soc.* **2009**, *131*, 12801–12808.
(34) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 9528–9533.
(35) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120*, 3265–3266.
(36) Kim, T.-Y.; Thompson, M. S.; Reilly, J. P. Peptide Photodissociation at 157 Nm in a Linear Ion Trap Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1657–1665.
(37) Moon, J. H.; Shin, Y. S.; Cha, H. J.; Kim, M. S. Photodissociation at 193 Nm of Some Singly Protonated Peptides and Proteins with M/z 2000–9000 Using a Tandem Time-of-Flight Mass Spectrometer Equipped with a Second Source for Delayed Extraction/post-Acceleration of Product Ions. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 359–368.
(38) Raspopov, S. A.; El-Faramawy, A.; Thomson, B. A.; Siu, K. W. M. Infrared Multiphoton Dissociation in Quadrupole Time-of-Flight Mass Spectrometry: Top-Down Characterization of Proteins. *Anal. Chem.* **2006**, *78*, 4572–4577.

(39) Guan, Z.; Kelleher, N. L.; O'Connor, P. B.; Aaserud, D. J.; Little, D. P.; McLafferty, F. W. 193 Nm Photodissociation of Larger Multiply-Charged Biomolecules. *Int. J. Mass Spectrom. Ion Process.* **1996**, *157–158*, 357–364.
(40) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**.
(41) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. Structural Characterization of a Partly Folded Apomyoglobin Intermediate. *Science* **1990**, *249*, 1544–1548.
(42) Eliezer, D.; Yao, J.; Dyson, H. J.; Wright, P. E. Structural and Dynamic Characterization of Partially Folded States of Apomyoglobin and Implications for Protein Folding. *Nat. Struct. Mol. Biol.* **1998**, *5*, 148–155.
(43) Lin, X.; Zhao, W.; Wang, X. Characterization of Conformational Changes and Noncovalent Complexes of Myoglobin by Electrospray Ionization Mass Spectrometry, Circular Dichroism and Fluorescence Spectroscopy. *J. Mass Spectrom.* **2010**, *45*, 618–626.
(44) Simmons, D. A.; Dunn, S. D.; Konermann, L. Conformational Dynamics of Partially Denatured Myoglobin Studied by Time-Resolved Electrospray Mass Spectrometry with Online Hydrogen-Deuterium Exchange. *Biochemistry***2003**, *42*, 5896–5905.
(45) Johnson, R. S.; Walsh, K. A. Mass Spectrometric Measurement of Protein Amide Hydrogen Exchange Rates of Apo- and Holo-Myoglobin. *Protein Sci.* **1994**, *3*, 2411–2418.
(46) Rouleau, M.; Patel, A.; Hendzel, M. J.; Kaufmann, S. H.; Poirier, G. G. PARP Inhibition: PARP1 and beyond. *Nat. Rev. Cancer* **2010**, *10*, 293–301.
(47) Langelier, M.-F.; Planck, J. L.; Roy, S.; Pascal, J. M. Structural Basis for DNA Damage–Dependent Poly(ADP-Ribosyl)ation by Human PARP-1. *Science* **2012**, *336*, 728–732.
(48) Tao, Z.; Gao, P.; Hoffman, D. W.; Liu, H. Domain C of Human Poly(ADP-Ribose) Polymerase-1 Is Important for Enzyme Activity and Contains a Novel Zinc-Ribbon Motif†,‡. *Biochemistry* **2008**, *47*, 5804–5813.
(49) Langelier, M.-F.; Servent, K. M.; Rogers, E. E.; Pascal, J. M. A Third Zinc-Binding Domain of Human Poly(ADP-Ribose) Polymerase-1 Coordinates DNA-Dependent Enzyme Activation. *J. Biol. Chem.* **2008**, *283*, 4105–4114.
(50) Li, H.; Robertson, A. D.; Jensen, J. H. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins Struct. Funct. Bioinforma.* **2005**, *61*, 704–721.
(51) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
(52) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.
(53) Fraczkiewicz, R.; Braun, W. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comput. Chem.* **1998**, *19*, 319–333.
(54) Shanmugam, G.; Selvi, C. C.; Mandal, A. B. Ethanol and Acetonitrile Induces Conformational Changes in Porcine Pepsin at Alkaline Denatured State. *Int. J. Biol. Macromol.* **2012**, *51*, 590–596.
(55) Sen, P.; Fatima, S.; Khan, J. M.; Khan, R. H. How Methyl Cyanide Induces Aggregation in All-Alpha Proteins: A Case Study in Four Albumins. *Int. J. Biol. Macromol.* **2009**, *44*, 163–169.
(56) Sen, P.; Iqbal, M. A.; Fatima, S.; Khan, R. H. Methyl Cyanide Induces A to B Transition and Aggregation at High Concentrations in E-State of Human Serum Albumin. *Biochem. Mosc.* **2010**, *75*, 367–374.
(57) Vahidi, S.; Stocks, B. B.; Liaghati-Mobarhan, Y.; Konermann, L. Submillisecond Protein Folding Events Monitored by Rapid Mixing and Mass Spectrometry-Based Oxidative Labeling. *Anal. Chem.* **2013**, *85*, 8618–8625.

(58) Konermann, L.; Vahidi, S.; Sowole, M. A. Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Anal. Chem.* **2014**, *86*, 213–232.
(59) O'Brien, J. P.; Mayberry, L. K.; Murphy, P. A.; Browning, K. S.; Brodbelt, J. S. Evaluating the Conformation and Binding Interface of Cap-Binding Proteins and Complexes via Ultraviolet Photodissociation Mass Spectrometry. *J. Proteome Res.* **2013**, *12*, 5867–5877.
(60) Cammarata, M.; Lin, K.-Y.; Pruet, J.; Liu, H.-W.; Brodbelt, J. Probing the Unfolding of Myoglobin and Domain C of PARP-1 with Covalent Labeling and Top-down Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 2534–2542.
(61) Heck, A. J. R. Native Mass Spectrometry: A Bridge between Interactomics and Structural Biology. *Nat. Methods* **2008**, *5*, 927–933.
(62) Belov, M. E.; Damoc, E.; Denisov, E.; Compton, P. D.; Horning, S.; Makarov, A. A.; Kelleher, N. L. From Protein Complexes to Subunit Backbone Fragments: A Multi-Stage Approach to Native Mass Spectrometry. *Anal. Chem.* **2013**, *85*, 11163–11173.
(63) Breuker, K.; McLafferty, F. W. Stepwise Evolution of Protein Native Structure with Electrospray into the Gas Phase, 10−12 to 102 S. *Proc. Natl. Acad. Sci.* **2008**, *105*, 18145–18152.
(64) Sharon, M.; Robinson, C. V. The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annu. Rev. Biochem.* **2007**, *76*, 167–193.
(65) Wood, T. D.; Chorush, R. A.; Wampler, F. M.; Little, D. P.; O'Connor, P. B.; McLafferty, F. W. Gas-Phase Folding and Unfolding of Cytochrome c Cations. *Proc. Natl. Acad. Sci.* **1995**, *92*, 2451–2454.
(66) Freitas, M. A.; Hendrickson, C. L.; Emmett, M. R.; Marshall, A. G. Gas-Phase Bovine Ubiquitin Cation Conformations Resolved by Gas-Phase Hydrogen/deuterium Exchange Rate and Extent. *Int. J. Mass Spectrom.* **1999**, *185–187*, 565–575.
(67) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. Protein Structure in Vacuo: Gas-Phase Conformations of BPTI and Cytochrome c. *J. Am. Chem. Soc.* **1997**, *119*, 2240–2248.
(68) Scarff, C. A.; Thalassinos, K.; Hilton, G. R.; Scrivens, J. H. Travelling Wave Ion Mobility Mass Spectrometry Studies of Protein Structure: Biological Significance and Comparison with X-Ray Crystallography and Nuclear Magnetic Resonance Spectroscopy Measurements. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3297–3304.
(69) Uetrecht, C.; Rose, R. J.; Duijn, E. van; Lorenzen, K.; Heck, A. J. R. Ion Mobility Mass Spectrometry of Proteins and Protein Assemblies. *Chem. Soc. Rev.* **2010**, *39*, 1633–1655.
(70) Lermyte, F.; Konijnenberg, A.; Williams, J. P.; Brown, J. M.; Valkenborg, D.; Sobott, F. ETD Allows for Native Surface Mapping of a 150 kDa Noncovalent Complex on a Commercial Q-TWIMS-TOF Instrument. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 343–350.
(71) Zhang, Z.; Browne, S. J.; Vachet, R. W. Exploring Salt Bridge Structures of Gas-Phase Protein Ions Using Multiple Stages of Electron Transfer and Collision Induced Dissociation. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 604–613.
(72) Zhang, H.; Cui, W.; Wen, J.; Blankenship, R. E.; Gross, M. L. Native Electrospray and Electron-Capture Dissociation FTICR Mass Spectrometry for Top-down Studies of Protein Assemblies. *Anal. Chem.* **2011**, *83*, 5598–5606.
(73) Zhang, H.; Cui, W.; Gross, M. L. Native Electrospray Ionization and Electron-Capture Dissociation for Comparison of Protein Structure in Solution and the Gas Phase. *Int. J. Mass Spectrom.* **2013**, *354–355*, 288–291.
(74) Breuker, K.; Brüschweiler, S.; Tollinger, M. Electrostatic Stabilization of a Native Protein Structure in the Gas Phase. *Angew. Chem. Int. Ed.* **2011**, *50*, 873–877.
(75) Schennach, M.; Breuker, K. Proteins with Highly Similar Native Folds Can Show Vastly Dissimilar Folding Behavior When Desolvated. *Angew. Chem. Int. Ed.* **2014**, *53*, 164–168.
(76) Li, H.; Wongkongkathep, P.; Orden, S. L. V.; Loo, R. R. O.; Loo, J. A. Revealing Ligand Binding Sites and Quantifying Subunit Variants of Noncovalent Protein Complexes in a Single Native Top-Down FTICR MS Experiment. *J. Am. Soc. Mass Spectrom.* **2014**, 1–9.

(77) Breuker, K.; Oh, H.; Horn, D. M.; Cerda, B. A.; McLafferty, F. W. Detailed Unfolding and Folding of Gaseous Ubiquitin Ions Characterized by Electron Capture Dissociation. *J. Am. Chem. Soc.* **2002**, *124*, 6407–6420.
(78) Li, H.; Wolff, J. J.; Van Orden, S. L.; Loo, J. A. Native Top-Down Electrospray Ionization-Mass Spectrometry of 158 kDa Protein Complex by High-Resolution Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2014**, *86*, 317–320.
(79) Skinner, O. S.; McLafferty, F. W.; Breuker, K. How Ubiquitin Unfolds after Transfer into the Gas Phase. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1011–1014.
(80) Robinson, E. W.; Leib, R. D.; Williams, E. R. The Role of Conformation on Electron Capture Dissociation of Ubiquitin. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1470–1480.
(81) Oh, H.; Breuker, K.; Sze, S. K.; Ge, Y.; Carpenter, B. K.; McLafferty, F. W. Secondary and Tertiary Structures of Gaseous Protein Ions Characterized by Electron Capture Dissociation Mass Spectrometry and Photofragment Spectroscopy. *Proc. Natl. Acad. Sci.* **2002**, *99*, 15863–15868.
(82) O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136*, 12920–12928.
(83) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86*, 2185–2192.
(84) Cannon, J. R.; Kluwe, C.; Ellington, A.; Brodbelt, J. S. Characterization of Green Fluorescent Proteins by 193 Nm Ultraviolet Photodissociation Mass Spectrometry. *PROTEOMICS* **2014**, *14*, 1165–1173.
(85) Catherman, A. D.; Durbin, K. R.; Ahlf, D. R.; Early, B. P.; Fellers, R. T.; Tran, J. C.; Thomas, P. M.; Kelleher, N. L. Large-Scale Top-down Proteomics of the Human Proteome: Membrane Proteins, Mitochondria, and Senescence. *Mol. Cell. Proteomics* **2013**, *12*, 3465–3473.
(86) Skinner, O. S.; Catherman, A. D.; Early, B. P.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L. Fragmentation of Integral Membrane Proteins in the Gas Phase. *Anal. Chem.* **2014**.
(87) Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662–666.
(88) Schwarzinger, S.; Mohana-Borges, R.; Kroon, G. J. A.; Dyson, H. J.; Wright, P. E. Structural Characterization of Partially Folded Intermediates of Apomyoglobin H64F. *Protein Sci.* **2008**, *17*, 313–321.
(89) Dasmeh, P.; Kepp, K. P. Unfolding Simulations of Holomyoglobin from Four Mammals: Identification of Intermediates and B-Sheet Formation from Partially Unfolded States. *PLoS ONE* **2013**, *8*, e80308.
(90) Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. Hydrogen/Deuterium Exchange Mass Spectrometry with Top-Down Electron Capture Dissociation for Characterizing Structural Transitions of a 17 kDa Protein. *J. Am. Chem. Soc.* **2009**, *131*, 12801–12808.
(91) Vahidi, S.; Stocks, B. B.; Konermann, L. Partially Disordered Proteins Studied by Ion Mobility-Mass Spectrometry: Implications for the Preservation of Solution Phase Structure in the Gas Phase. *Anal. Chem.* **2013**, *85*, 10471–10478.
(92) Teale, F. W. Cleavage of the Haem-Protein Link by Acid Methylethylketone. *Biochim. Biophys. Acta* **1959**, *35*, 543.
(93) Reymond, M. T.; Dyson, H. J.; Wright, P. E.; Merutka, G. Folding Propensities of Peptide Fragments of Myoglobin. *Protein Sci.* **1997**, *6*, 706–716.
(94) Fraczkiewicz, R.; Braun, W. Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules. *J. Comput. Chem.* **1998**, *19*, 319–333.
(95) Lin, X.; Zhao, W.; Wang, X. Characterization of Conformational Changes and Noncovalent Complexes of Myoglobin by Electrospray Ionization Mass Spectrometry, Circular Dichroism and Fluorescence Spectroscopy. *J. Mass Spectrom.* **2010**, *45*, 618–626.

(96) Tirado-Rives, J.; Jorgensen, W. L. Molecular Dynamics Simulations of the Unfolding of Apomyoglobin in Water. *Biochemistry* **1993**, *32*, 4175–4184.
(97) Ly, T.; Julian, R. R. Elucidating the Tertiary Structure of Protein Ions in Vacuo with Site Specific Photoinitiated Radical Reactions. *J. Am. Chem. Soc.* **2010**, *132*, 8602–8609.
(98) Whitaker, T. L.; Berry, M. B.; Ho, E. L.; Hargrove, M. S.; Phillips, G. N.; Komiyama, N. H.; Nagai, K.; Olson, J. S. The D-Helix in Myoglobin and in the Beta Subunit of Hemoglobin Is Required for the Retention of Heme. *Biochemistry* **1995**, *34*, 8221–8226.
(99) Brunet, C.; Antoine, R.; Dugourd, P.; Canon, F.; Giuliani, A.; Nahon, L. Photo-Induced Electron Detachment of Protein Polyanions in the VUV Range. *J. Chem. Phys.* **2013**, *138*, 064301.
(100) Chien, Y.-T.; Hwang, J.-K.; Huang, S.-W. On the Relationship Between Residue Solvent Exposure and Thermal Fluctuations in Proteins. In *Protein Structure*; Faraggi, E., Ed.; InTech, 2012.
(101) Zhang, H.; Zhang, T.; Chen, K.; Shen, S.; Ruan, J.; Kurgan, L. On the Relation between Residue Flexibility and Local Solvent Accessibility in Proteins. *Proteins Struct. Funct. Bioinforma.* **2009**, *76*, 617–636.
(102) Tsai, C. S. *Biomacromolecules: Introduction to Structure, Function and Informatics*; John Wiley & Sons, 2007.
(103) Eliezer, D.; Wright, P. E. Is Apomyoglobin a Molten Globule? Structural Characterization by NMR. *J. Mol. Biol.* **1996**, *263*, 531–538.
(104) Nishimura, C.; Dyson, H. J.; Wright, P. E. Identification of Native and Non-Native Structure in Kinetic Folding Intermediates of Apomyoglobin. *J. Mol. Biol.* **2006**, *355*, 139–156.
(105) Correcirc, D. H. A.; a; Ramos, C. H. I. The Use of Circular Dichroism Spectroscopy to Study Protein Folding, Form and Function. *Afr. J. Biochem. Res.* **2009**, *3*, 164–173.
(106) Kawamura-Konishi, Y.; Kihara, H.; Suzuki, H. Reconstitution of Myoglobin from Apoprotein and Heme, Monitored by Stopped-Flow Absorption, Fluorescence and Circular Dichroism. *Eur. J. Biochem.* **1988**, *170*, 589–595.
(107) Wang, F.; Tang, X. Conformational Heterogeneity and Stability of Apomyoglobin Studied by Hydrogen/Deuterium Exchange and Electrospray Ionization Mass Spectrometry†. *Biochemistry* **1996**, *35*, 4069–4078.